DOCUMENT RESUME

ED 219 033 HE 015 303

AUTHOR Hendel, Patricia K., Ed.
TITLE Abstracts from the Research and Development Report Series, 1963-81. College Board Report No. 82-1.
INSTITUTION College Entrance Examination Board, New York, N.Y.
PUB DATE 82
NOTE 52p.
AVAILABLE FROM College Board Publication Orders, Box 2815, Princeton, NJ 08541 ($4.00).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Academic Achievement; Achievement Tests; Aptitude Tests; Attitude Measures; College Admission; *College Entrance Examinations; Educational Development; *Educational Research; Educational Testing; Followup Studies; Higher Education; Institutional Research; Predictive Measurement; Questionnaires; Research Design; *Research Methodology; Research Projects; *Standardized Tests; *Student Evaluation; Surveys; *Testing
IDENTIFIERS *College Entrance Examination Board

ABSTRACT

A compilation of 109 abstracts from the "Research and Development Report Series," is presented based on R&D projects conducted during 1963-1981 by Educational Testing Service on behalf of the College Entrance Examination Board. Project titles and authors include the following: "The Effect of Special Preparation on SAT Verbal Scores" (Donald L. Alderman, Donald E. Powers); "The Development of Statistical Indices for Detecting Cheaters" (William H. Angoff); "Anchored Scaling and Equating: Old Conceptual Problems and New Methods" (Robert F. Boldt); "Group Comparisons for the Test of Standard Written English" (Hunter M. Breland); "Validation of the Descriptive Tests of Mathematics Skills" (Brent Bridgemen); "Testing of Culturally Different Groups" (Joel Campbell); "A Method for Comparing the Performance of Different Groups on the Items in a Test" (Carolyn Cardall, William E. Coffman); "College Decisions on Advanced Placement" (Patricia L. Casserly, Richard E. Peterson, William E. Coffman); "Black Students at Predominantly White Colleges: A Research Description" (John A. Centra); "Test Bias: Validity of the Scholastic Aptitude Test for Negro and White Students in Integrated Colleges" (T. Anne Cleary); "Formulating Hypotheses: Factor Structure and Validity Data" (Rosemary Cliff and Normal Cliff); "A Factor Analysis of the Verbal Sections of the Scholastic Aptitude Test" (William E. Coffman); "Faculty Perceptions of Students" (Junius A. Davis); "Colleges' Use and Evaluation of the CEEB Writing Sample" (Ruth B. Ekstrom); "A Study of the Relationships Among Speed and Power, Aptitude Test Scores, and Ethnic Identity" (Franklin R. Evans). (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219033

Abstracts from the Research and Development Report Series, 1963-81

Edited by
Patricia K. Hendel

College Board Report No. 82-1

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

CEEB

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 015 303

College Entrance Examination Board, New York, 1982

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publication Orders, Box 2815, Princeton, New Jersey 08541. The price is $4.

Copyright © 1982 by College Entrance Examination Board.
All rights reserved.
Printed in the United States of America.

CONTENTS

Preface ix

The Effect of Special Preparation on SAT Verbal Scores, D. L. Alderman and D. E. Powers 1

The Development of Statistical Indices for Detecting Cheaters, W. H. Angoff 1

Item-Race Interaction on a Test of Scholastic Aptitude, W. H. Angoff and S. F. Ford 1

Equating the Scales of the Spanish-Language Prueba de Aptitud Academica and the English-Language Scholastic Aptitude Test of the College Entrance Examination Board, W. H. Angoff and C. C. Modu 1

Patterns of Test and Item Difficulty for Six Foreign Language Groups on the Test of English as a Foreign Language, W. H. Angoff and A. T. Sharon 2

The Equating of the Scales for the Canadian and American Scholastic Aptitude Tests, W. H. Angoff and J. Stern 2

Anchored Scaling and Equating: Old Conceptual Problems and New Methods, R. F. Boldt 3

Concurrent Validity of the PAA and SAT for Bilingual Dade County High School Volunteers, R. F. Boldt 3

A Study of Linearity and Homoscedasticity of Test Scores in the Chance Range, R. F. Boldt 3

Group Comparisons for the Test of Standard Written English, H. M. Breland 3

A Study of College English Placement and the Test of Standard Written English, H. M. Breland 4

Population Validity and College Entrance Measures, H. M. Breland and S. Minsky . . . 4

Validation of the Descriptive Tests of Mathematics Skills, B. Bridgeman with a contribution by R. G. Kline 5

Testing of Culturally Different Groups, J. Campbell 5

A Method for Comparing the Performance of Different Groups on the Items in a Test, C. Cardall and W. E. Coffman 5

College Decisions on Advanced Placement. I. A Follow-up of Advanced Placement Candidates of 1963, P. L. Casserly 6

College Level Examination Program: Its Meaning to Participants, P. L. Casserly . . . 6

To See Ourselves as Others See Us: An Evaluation of the Advanced Placement Program by Students from 252 Secondary Schools at 20 Colleges, P. L. Casserly . . . 7

College Decisions on Advanced Placement. II. A Report of Interviews Concerned with Advanced Placement Policies and Practices at 63 Colleges and Universities, P. L. Casserly and R. E. Peterson 7

College Decisions on Advanced Placement. II. An Interview Survey of Advanced Placement Policies at 63 Colleges, P. L. Casserly and R. E. Peterson 7

Black Students at Predominantly White Colleges: A Research Description, J. A. Centra 7

College Enrollment in the 1980s: Projections and Possibilities, J. A. Centra 7

The College Environment Revisited: Current Descriptions and a Comparison of Three Methods of Assessment, J. A. Centra 8

Test Bias: Validity of the Scholastic Aptitude Test for Negro and White Students in Integrated Colleges, T. A. Cleary 8

An Investigation of Item Bias, T. A. Cleary and T. L. Hilton 9

Formulating Hypotheses: Factor Structure and Validity Data, R. Cliff and N. Cliff 9

A Factor Analysis of the Verbal Sections of the Scholastic Aptitude Test, W. E. Coffman 9

On the Validity of Essay Tests of Achievement, W. E. Coffman 9

A Report on the Pretesting for the Scholastic Aptitude Test in Spanish-Form APR, W. E. Coffman 10

The Scholastic Aptitude Test of 1963-64, W. E. Coffman 10

Single Score Versus Multiple Score Reading of the American History Advanced Placement Examination, W. E. Coffman and D. G. Kurfman 10

The Relationship Between Scores on the Scholastic Aptitude Test and Certain Post-College Activities of Phi Beta Kappa Members of the Yale Classes of 1931-1950, W. E. Coffman and M. H. Mahoney 10

Effects of an Accelerated Reading Course on SAT-V Scores, W. E. Coffman and M. E. Neun 11

Faculty Perceptions of Students, I. The Development of the Student Rating Form, J. A. Davis 11

Faculty Perceptions of Students, II. Faculty Definition of Desirable Student Traits, J. A. Davis 11

Faculty Perceptions of Students, III. Structure of Faculty Characterizations, J. A. Davis 11

Faculty Perceptions of Students, IV. Desirability and Perception of Academic Performance, J. A. Davis 12

Faculty Perceptions of Students, V. A Second-Order Structure for Faculty Characterizations, J. A. Davis 12

Faculty Perceptions of Students, VI. Characteristics of Students for Whom There Is Faculty Agreement on Desirability, J. A. Davis 12

Colleges' Use and Evaluation of the CEEB Writing Sample, R. B. Ekstrom 12

A Study of the Relationships Among Speed and Power, Aptitude Test Scores, and Ethnic Identity, F. R. Evans 13

Project Access Research Report No. 2, Patterns of Test Performance by High School Students of Four Ethnic Identities, R. L. Flaugher 13

Project Access Research Report No. 3, Minority Versus Majority Group Performance on an Aptitude Test Battery, R. L. Flaugher 13

Reactions to a Very Difficult Test by an Inner-City High School Population: A Test and Item Analysis, R. L. Flaugher and L. W. Pike 14

Project Access Research Report No. 5, Patterns of Ability Factors Among Four Ethnic Groups, R. L. Flaugher and D. A. Rock 14

The Wide Range Validity of Certain New Aptitude Tests, R. L. Flaugher and D. A. Rock 14

The College Board Scholastic Aptitude Test as a Predictor of Academic Achievement in Secondary Schools in England, J. Fremer, W. E. Coffman, and P. H. Taylor 14

The Revision of the Interest Index, J. French 15

Project Access Research Report No. 4, Influence of Race, Sex, and City on Inductive Reasoning Items, A. H. Gitlitz and N. L. Kaufman 15

Studies of the Question Types in the CEEB English Composition Test as Predictors of an Essay Criterion, F. I. Godshalk, F. Swineford, and W. E. Coffman 15

An Investigation into the Construct Validity of Three Career Skills Exercises, J. E. Grandy 16

Scale Properties of the Interest Index, G. Halpern 16

Effectiveness of PSAT and SAT Booklets, I. Bulletin for Students, 1964 Preliminary Scholastic Aptitude Test, G. Halpern and M. Sasajima 16

Effectiveness of PSAT and SAT Booklets, II. Your College Board Scores: Preliminary Scholastic Aptitude Test, G. Halpern and M. Sasajima 16

Effectiveness of PSAT and SAT Booklets, III. A Description of the College Board Scholastic Aptitude Test, G. Halpern and M. Sasajima 17

Effectiveness of PSAT and SAT Booklets, IV. Your College Board Scores: Scholastic Aptitude Test, Achievement Tests, G. Halpern and M. Sasajima 17

A Summary of Current Social Studies Curriculum Projects, S. E. Harrison 17

Learning Style Preferences Among College Students, R. T. Hartnett 17

Financial Aid to College Students, 1963-64, E. W. Haven and R. E. Smith 18

A Survey of College Physics Departments Regarding Advanced Placement Practices, L. H. Heidgerd 18

Growth Study, II. Personal Background, Experience, and School Achievement: An Investigation of the Contribution of Questionnaire Data to Academic Prediction, T. L. Hilton and A. E. Myers 18

Estimating Parents' Contribution to College Costs: The Accuracy of Three Measures of Succeeding Year Family Net Income, D. H. Horch and A. T. Sharon 19

College Board Scores of Candidates of Non-English-Language Background Tested at Foreign Centers, J. J. Howell 19

Language Background of Candidates for College Board Examinations Tested at Foreign Centers, J. J. Howell 20

Verbal and Mathematical Ability of High School Juniors in 1974: A Norms Study of PSAT/NMSQT, R. Jackson and W. B. Schrader 20

Effects of Coaching on the College Board English Composition Test, P. I. Jacobs . . . 21

College Mathematics Courses Through the Calculus: A Survey, C. O. Jones and S. I. Williams 21

Change in Scholastic Aptitude Test Scores of Candidates Tested in the Junior and Senior Years 1961-1964, S. F. Karas 21

Development of a Reduced Set of Composite Equations for Three Predictors, S. F. Karas and L. M. Kendall 22

First-Year College Biology Courses: A Survey, W. Kastrinos and G. Sanders 22

The Measurement of Academic Interests, Part I. Characteristics of the Academic Interest Measures, M. R. Katz, L. Norris, and G. Halpern 22

The CEEB English Language-Study Test: A Survey of Linguistics Units in the High Schools, M. K. Levin 22

Measuring the Appropriateness of Multiple-Choice Test Scores, M. V. Levine and D. B. Rubin 23

Grade Adjustments for Prediction of Academic Performance: A Review, R. L. Linn . . . 23

Correlates of Academic Performance of Community College Students in Career or Transfer Programs: A Pilot Study, R. L. Linn and J. A. Davis 23

Sequential Testing for Dichotomous Decisions, R. L. Linn, D. A. Rock, and T. A. Cleary 23

Growth Study I. Adolescent Behavior and Interests, M. H. Maier and S. B. Anderson . 23

The Characteristics of Preliminary Scholastic Aptitude Test User Schools and Non-User Schools, R. M. Majetic 24

Item Characteristic Curve Solutions to Three Intractable Testing Problems, G. L. Marco 24

The Study of Change in CEEB Achievement Tests Over a Ten-Year Period, D. Mills . . . 24

A Description of the Satisfaction Questionnaire for Junior Colleges in Terms of Rotated Factors, C. C. Modu 24

The Effectiveness of an Essay Section in the American History and Social Studies Test, C. C. Modu 25

The Stability of the SAT Score Scale, C. C. Modu and J. Stern 25

A Survey of College Biology Departments Regarding Introductory Course Curricula and Advanced Placement Practices, M. R. Moore 25

Reliability of Reading of the College Board English Composition Test, December 1963, A. E. Myers, C. B. McConville, and W. E. Coffman 25

A Comparison of Three Methods for Scoring Multiple-Choice Tests, C. T. Myers 26

The Measurement of Academic Interests, Part II. The Predictive Validities of Academic Interest Measures, L. Norris and M. R. Katz 26

Short-Term Instruction, Testwiseness, and the Scholastic Aptitude Test: A Literature Review with Research Recommendations, L. W. Pike 27

The Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items, L. W. Pike and F. R. Evans 27

Follow-up Study of a National Sample of High School Seniors: Ability Level of Students Who Entered College Directly from High School and Were Graduated from College Within Four Years, B. Pitcher 27

A Preliminary Study of the Prediction of College Freshman English Grades from College Board Achievement Tests in the Foreign Languages, B. Pitcher and J. Selsam 28

Effects of "Taking the SAT" on Rate of Work and Guessing, D. E. Powers 28

Students' Use of and Reactions to Alternative Methods of Preparing for the SAT, D. E. Powers 29

The Use, Acceptance, and Impact of "Taking the SAT"--A Test Familiarization Booklet, D. E. Powers and D. L. Alderman 29

A Group Approach to the Analysis of Individual Differences in the Randomness of Guessing Behavior on Multiple-Choice Tests and the Development of Scoring Methods to Take Such Differences into Account, D. B. Price 29

A Factor Analysis of the Mathematical Sections of the Scholastic Aptitude Test, R. M. Pruzek and W. E. Coffman 30

Handicapped Students and the SAT, M. Ragosta 30

The Effect of Special Instruction Upon Test Performance of High School Students in Tennessee, S. O. Roberts and D. B. Oppenheim 31

Construct Validity of the SAT Across Populations--An Empirical Confirmatory Study, D. A. Rock and C. E. Werts 31

Interaction Between College Effects and Students' Aptitudes, D. A. Rock, L. L. Baird, and R. L. Linn 31

High School Curriculum and Performance on Mathematics Achievement Tests, D. C. Ross 31

Financial Aid Awards of High School Seniors in the Northeastern States as a Function of Socioeconomic Class, G. A. Schlekat 32

Follow-up Study of a National Sample of High School Seniors: Phase 2--One Year After Graduation, D. W. Seibel 32

A Study of the CGP Financial Need Indicator, L. Sequeira 32

Three Methods of Testing Literary Comprehension at the Advanced Placement Level, A. N. Serling 33

Effectiveness of Remediation in Junior College, A. T. Sharon 33

Measurement of College Achievement by the College Level Examination Program, A. T. Sharon 33

Using Scholastic Aptitude Test Scores in Counseling Prospective Graduate and Law Students, Statistical Analysis Division 33

The Stability of the SAT Verbal Score Scale, E. E. Stewart 34

A Comparison of the Performance of BSCS and Non-BSCS Candidates on the CEEB Biology Achievement Test May 1963 Administration, D. W. Stickell 34

An Analysis of Culver Military Academy Data Relating to Performance on the College Board Chemistry Test as Affected by Curricular Change, D. W. Stickell . . . 35

Test Bias: Validity of the SAT for Blacks and Whites in Thirteen Integrated Institutions, G. Temp 35

A Study of the Comparative Predictive Validities of the Essay and Objective Sections of the College Entrance Examination Board Advanced Placement Examination in Physics, R. E. Thompson 35

Developing a Collegiate Physical Science Test, R. E. Thompson 36

Investigations of the Appropriateness of the College Board Science Achievement Tests for Students of Different High School Science Courses, R. E. Thompson 36

Prediction of College Achievement Among Mexican-American Students in California, J. R. Warren 37

Student Reactions to College: The Development of a Questionnaire Through Which Junior College Students Describe Their College Experiences, J. R. Warren and P. J. Roelfs 37

The Usefulness of Adjustments to Secondary School Grades in the Prediction of College Success, R. W. Watkins and R. S. Levine 37

Multivariate Talent Flow Analysis: A Pilot Study, K. M. Wilson 38

Predicting the Long-Term Performance in College of Minority and Nonminority Students: A Comparative Analysis in Two Collegiate Settings, K. M. Wilson 39

Index of Authors 41

PREFACE

Much of the research and development program of the College Board is conducted by Educational Testing Service in Princeton, New Jersey. During the period 1963 through early 1981, the results of 120 R&D projects conducted by ETS on behalf of the College Board were published in the Research and Development Report (RDR) series. That series ended with the introduction in 1981 of the College Board Report series, which will include ETS studies on behalf of the Board as well as other reports.

The compilation of abstracts from the RDR series that follows is based primarily on a collection of such abstracts by Linda Johnson in the ETS Office of Program Research. One hundred and nine abstracts are included; abstracts of the remaining eleven studies are not available.

Studies included in the RDR series, and now in the College Board Report series, have been published in limited quantities. As a result, many of the RDR reports are out of print. Reports in the College Board Report series are listed in the College Board Publications Catalog, which may be obtained by writing Box 2815, Princeton, New Jersey 08541. Individuals seeking further information about reports that are out of print should write the Coordinator of Research and Development Planning, The College Board, 888 Seventh Avenue, New York, New York 10106.

ALDERMAN, DONALD L., AND DONALD E. POWERS. The Effect of Special Preparation on SAT Verbal Scores (RDR 78-79, No. 4). February 1979.

This study examined the effectiveness of secondary school programs specifically designed to improve student performance on the SAT verbal component. Students at eight schools were randomly assigned to either a treatment group which was given special preparation for the SAT-V or a control group with access to the same preparation delayed for the purposes of the study. The programs of special preparation were those already in place at the public and private secondary schools selected as participants. A special administration of the SAT served as the posttest. Special preparation resulted in an overall difference of eight points, on the SAT-V scale of 200-800 points, between the treatment and control groups. The difference corresponded to one additional item correct for students receiving special preparation and stemmed primarily from student performance on analogy and antonym test items.

ANGOFF, WILLIAM H. The Development of Statistical Indices for Detecting Cheaters (RDR 72-73, No. 1). July 1972.

Comparison data on SAT verbal and mathematical were collected on pairs of examinees in three samples for later use in detecting instances of willful copying. Two of the samples were constructed with the knowledge that no examinees could possibly have copied from the answer sheet of any other examinee in the sample. The third sample was taken entirely from a single center believed to be free of cheating. In each sample the answer sheet of each examinee was compared with the answer sheet of every other examinee. Eight detection indices were developed and distributions were run for possible operational use in making future judgments regarding examinees who were actually suspected of copying. Covariance analyses between samples indicated statistical but not practical significance, and consequently it was judged that any one of the samples could serve the purposes of operational detection as well as either of the other two.

Empirical tryout of the indices against known and admitted copiers gave some results that permitted the elimination of three of the indices from further use. Practical considerations removed a fourth, and further statistical study eliminated two others. The remaining two have been in successful operational use at Educational Testing Service for more than two years.

ANGOFF, WILLIAM H., AND SUSAN F. FORD. Item-Race Interaction on a Test of Scholastic Aptitude (RDR 71-72, No. 3). October 1971.

Several samples of Black and white students were drawn from the 1970 PSAT administration in Georgia and studied for item x race interaction on both the verbal and mathematical sections of the test. When subsamples of candidates were drawn from their respective racial groups, matched on mathematical for the study of verbal items, and matched on verbal for the study of mathematical items, there was an observable decrease in the size of the item x race interaction, suggesting that one factor contributing to that interaction was simply the difference in performance levels on the test shown by the two races. Further analyses demonstrated a moderate item x group interaction for Blacks native to different cities and a moderate item x group interaction for Blacks native to areas of different population density.

ANGOFF, WILLIAM H., AND CHRISTOPHER C. MODU. Equating the Scales of the Spanish-Language Prueba de Aptitud Academica and the English-Language Scholastic Aptitude Test of the College Entrance Examination Board (RDR 72-73, No. 4). January 1973.

The purpose of this study was to establish score equivalencies between the College Board Scholastic Aptitude Test (SAT) and its Spanish-language equivalent, the College Board Prueba de Aptitud Academica (PAA). The method of the study involved two phases: the selection of test items, equally appropriate and useful for Spanish- and English-speaking students, for use in the equating of the two tests, and the equating analysis itself. The method of the first phase was to choose two sets of items, one originally appearing in Spanish, the other originally appearing in English; to translate each set into the other language; and to administer both sets in the appropriate language mode

for pretest purposes to both types of students. These administrations were conducted in the fall of 1970 with samples of candidates taking the PAA or the SAT at regularly scheduled administrations. They provided data regarding the difficulty and discrimination power of each item for each of the two groups and, what was of special interest, an index of appropriateness of each item for both groups.

On the basis of the analyses of these data, two sets of items, one in verbal and the other in mathematical, were chosen and assembled as "common items" to be used for equating. In the second phase of the study, these "common items," appearing in Spanish and also in English, were administered in the appropriate language along with the operational form of the PAA in November, 1971, and with the operational form of the SAT in January, 1972. The data resulting from the administrations of these "common items" were used to calibrate for differences in the abilities of the two groups of candidates and permitted both linear and equipercentile equating of the two tests. Conversion tables relating the PAA verbal scores to the SAT verbal scores and the PAA mathematical scores to the SAT mathematical scores are given. These conversions represent an average of the linear and equipercentile results. Because of the scarcity of data at the upper end of the distribution of PAA scores, score equivalencies are permissible, strictly speaking, only as high as the mid-700s. Score equivalencies beyond the mid-700s were obtained by extrapolation.

ANGOFF, WILLIAM H., AND AMIEL T. SHARON. Patterns of Test and Item Difficulty for Six Foreign Language Groups on the Test of English as a Foreign Language (RDR 71-72, No. 5). February 1972.

Two-factor analyses of variance with multiple measurements on one factor were conducted for the five sections of the Test of English as a Foreign Language (TOEFL) and six language groups, and also for the items of each section, based on data collected at the first administration of this form of the test, in October, 1969. The principal findings, that the test x group interaction and the item x group interactions were significant, indicated that some tests and some items were relatively more difficult for some language groups than for others.

Displays of the item x group interactions were made by analyzing the item difficulty plots for each language group against a spaced sample of all candidates taking this form of the test in October, 1969. A measure of the deviation of each item from the central tendency of the plot was developed, expressing the degree to which the item was especially difficult--or easy--for a particular language group relative to the other items. Distributions of these measures are given for each of the five parts of TOEFL, for each of the six language groups.

ANGOFF, WILLIAM H., AND JUNE STERN. The Equating of the Scales for the Canadian and American Scholastic Aptitude Tests (RDR 71-72, No. 4). December 1971.

In this equating study the form of the CSAT administered in January, 1971, was equated to the form of the SAT also administered in January, 1971. The design for equating involved the administration of a one-half-hour operational verbal section and a one-half-hour operational mathematical section of the CSAT as "experimental" (nonoperational) tests to random samples of the January, 1971, SAT group. Similarly, a one-half-hour operational verbal section and a one-half-hour operational mathematical section of the SAT were administered as "experimental" (nonoperational) tests to random samples of the January, 1971, CSAT group. As a result of these symmetrical "crossover" administrations, random samples of the American and the Canadian groups took the same two sections of verbal test materials. Other random samples of the American and Canadian groups took the same two sections of mathematical test materials. These common sections, representing 95 verbal items and 45 mathematical items, were then item analyzed, and item difficulty (delta) plots were made to identify "outlier" items--items that were not equally appropriate for the two groups of examinees. As a result of these analyses, 20 verbal and 10 mathematical items were identified as "outliers" and removed from the equating sections. Data on new samples were developed with the remaining 75 verbal equating items and 35 mathematical equating items to equate raw scores on that form of the CSAT to corresponding scores on that form of the SAT. With these equations in hand and with the equations

converting each set of raw scores to its own reporting scale, conversion equations were developed relating the verbal and mathematical scales for the CSAT to the corresponding scales for the SAT.

BOLDT, ROBERT F. Anchored Scaling and Equating: Old Conceptual Problems and New Methods (RDR 72-73, No. 2). August 1972.

This paper describes several situations in which generalizations of statistical results are not possible by representative sampling but which are attempted using corrections for selection of groups. The situations include hiring, admissions, differential classification, guidance, test score equating, and test score scaling. Evidence of inaccuracies of the assumptions underlying the corrections is adduced. The Pearson equations, which rest on these assumptions, are mentioned as a basis of scaling and equating procedures in existence, an alternative approach is suggested, and its application to anchored equating, vertical equating, scaling, and equating with mixed essay and objective material is described. The alternative approach consists of a principle for choosing objective functions whose optimization would lead to a selection of conversion constants for equating. The principle is that equal equating test scores should be associated with equal reported scores on the average. Constrained optimizations are suggested where policy considerations so indicate.

BOLDT, ROBERT F. Concurrent Validity of the PAA and SAT for Bilingual Dade County High School Volunteers (RDR 68-69, No. 3). April 1969.

In order to serve prospective college students who have a Spanish-language background, the College Entrance Examination Board developed the Prueba de Aptitud Academica (PAA), an examination that assesses verbal and mathematical aptitude as does the Scholastic Aptitude Test, but does it in the Spanish language. The purposes of this study were (1) to develop and compare PAA-based and SAT-based equations for estimating the high school grade point average (GPA) and (2) to determine whether measures of the student's orientation toward the English or Spanish language (LO) or a measure of his socio-economic status (SES) could be used together with his SAT and PAA scores to gain improved concurrent validity.

It was found that for the group studied, if 200 points were subtracted from both PAA-V and PAA-M, the resulting variables could be used in an SAT formula for predicting GPA, though the validity would be about 15 correlation points lower.

No gain in validity was obtained by using SES or LO measures as moderators of the regressions of GPA on SAT and PAA. In fact, SAT-M alone did about as well as any of the moderation systems.

BOLDT, ROBERT F. A Study of Linearity and Homoscedasticity of Test Scores in the Chance Range (RDR 66-67, No. 6). August 1966.

Abstract not available.

BRELAND, HUNTER M. Group Comparisons for the Test of Standard Written English (RDR 77-78, No. 1). August 1977.

Group comparisons for male, female, majority, and minority students were conducted for the Test of Standard Written English (TSWE). Data for two academic years and from 18 different institutions were analyzed by pooling data across institutions within each of the academic years. Analyses of data from the first academic year focused on relationships between TSWE scores and grades. Analyses of the second academic year focused on relationships between TSWE scores and scores on graded writing samples. No important group differences in traditional correlation analyses for either grade or essay prediction were observed. No important group differences were observed in either correlation analyses or regression analyses of the second year essay data. Analyses of correct and incorrect placement decisions (hits and misses) at specific TSWE cutoff scores revealed no noteworthy group differences whether outcomes were based on English course grades or on freshman year writing performance. Whether grades or essays were used as the outcome, the proportion of incorrect decisions (misses) was less for minority students than for any of the other groups. For all groups, the TSWE appeared

to predict freshman year writing performance as well as or better than pre-course writing samples, high school English grades, or high school rank in class.

BRELAND, HUNTER M. A Study of College English Placement and the Test of Standard Written English (RDR 76-77, No. 4). January 1977.

As a supplement to previous studies and analyses involving larger numbers of institutions and students, four institutions collaborated with Educational Testing Service in a special study of college freshman English placement practices and a new placement test, the Test of Standard Written English (TSWE). The small number of institutions made possible analyses that had not been possible in previous large-scale studies. The institutions provided data on student performance during college, and Educational Testing Service matched these data with files of other data obtained at the time the students applied for college. There were four results of the study that are of principal interest. First, student performances on essay tests of writing ability were shown to have a strong relationship to student performances on multiple-choice tests of writing ability as represented by the TSWE. Second, the analyses indicated that a brief multiple-choice test of writing ability predicted actual writing performance during the freshman year as well as or better than a brief essay test given at the beginning of the freshman year. Third, analyses indicated that both the TSWE and the Scholastic Aptitude Test (SAT) contribute more to the prediction of freshman writing performance than either high school English grades or high school rank. Fourth, the results suggest that the TSWE is more useful in placing students in English composition courses than the SAT, even though a combination of the two provides a better prediction of writing performance than either used alone.

BRELAND, HUNTER M., AND SHULA MINSKY. Population Validity and College Entrance Measures (RDR 78-79, No. 2). November 1978.

This report reviews and summarizes the data of published reports and papers pertaining to the validity of measures commonly used at college entrance. These measures consist of various types of quantitative indices used to describe the high school record and various types of academic tests. The data were categorized by sex and ethnic classifications where possible and comparisons were made.

Most of the observations made in previous reviews of this kind were upheld. Both the high school record and standardized academic tests appear to be generally useful at college entrance. Both tend to predict college performance, but they are most appropriately used in combination. For Anglo populations, the high school record usually yields slightly higher predictive correlations. For Black populations (except Black females), however, the data suggest no superiority of the high school record as a predictor of college performance. The amount of data available for Chicano samples was not sufficient for any generalizations about predictive correlations, but the magnitude of those correlations available were on the average lower than those for Anglo populations. For recent years, correlations for Black samples tended to be slightly lower than those for Anglo samples.

A primary objective in the review was to consider whether inferences made from the high school record and test scores about different populations tended to be correct inferences. Where inferences tended to be systematically incorrect, there was an interest in the direction of the error. Since the high school record and test scores are used in making inferences about probable college performance, the systematic inferential errors of interest consisted of the degree to which college performance of populations were over- or underpredicted when the data used in making inference were not limited to data for that population.

Data from a number of studies show that the college performance of Black populations, male and female, have been consistently overpredicted by the traditional academic predictors (the high school record and standardized test scores) when predictions are based on data from white or predominantly white samples. The high school record would appear, from the studies reviewed, to be the principal source of this overprediction. Standardized test scores also have tended to overpredict college performance of Black populations, but a combination of the high school record and test scores appears to minimize overprediction. In contract,

published data show women to be consistently underpredicted by the traditional predictors when predictions are based on data from male or predominantly male samples.

This review indicates a number of topics where further research appears to be needed. These topics include the study of: long-range criteria, objective criteria, prediction within major-field, the high school record for different populations, college grades for different populations, neglected populations, factorial consistency of measures across populations, item characteristic curves, expert judgment of test content, and the rationale of the content of standardized tests.

BRIDGEMAN, BRENT, WITH A CONTRIBUTION BY ROBERTA G. KLINE. Validation of the Descriptive Tests of Mathematics Skills (RDR 79-80, No. 2). June 1980.

A variety of techniques were used to assess the validity of the Descriptive Tests of Mathematics Skills (DTMS) for making placement decisions. DTMS tests were administered in 36 institutions that represented a broad spectrum of two-year and four-year colleges. A pre-post design was used in seven institutions, with a pretest only in the other colleges.

Results indicated that the DTMS tests were: (1) sensitive to gains from instruction in a single semester; (2) highly correlated with course grades obtained concurrently; (3) predictive of end-of-course grades when the DTMS tests were administered at the beginning of the course; (4) a better predictor of course grades than SAT mathematical scores for remedial courses; (5) related to course difficulty as perceived by students; (6) capable of demonstrating a trait-treatment interaction in a precalculus-calculus sequence; and (7) related to the content of mathematics courses as judged by mathematics faculty members.

CAMPBELL, JOEL. Testing of Culturally Different Groups (RDR 63-64, No. 14). June 1964.

The purpose of this study was to summarize research findings on testing of individuals from groups, particularly Negroes, that differ in some respects from the majority culture in the United States. Particular attention was paid to factors that appeared to cause or be associated with changes in test scores. The effectiveness of tests in predicting performance of Negroes in integrated and segregated colleges was evaluated.

A program of research, designed to fill gaps in present knowledge, was outlined, with particular attention focused on selection for college and remedying educational deficits at the college level.

CARDALL, CAROLYN, AND WILLIAM E. COFFMAN. A Method for Comparing the Performance of Different Groups on the Items in a Test (RDR 64-65, No. 9). November 1964.

One of the persistent problems facing the researcher in the field of testing is that of determining the extent to which a test is appropriate for use with different subsamples of a general population. An obvious, but somewhat difficult approach to the problem would be to validate the test under consideration for every identifiable subgroup in the population and to develop specific interpretive guides for use with each of the subgroups. Such an approach, however, throws light on the question of what factors are responsible for any differences which appear. In order to develop an understanding, one must deal with the data at the level of the individual test item.

But a direct comparison of the difficulty of an individual item for subsamples of the population is complicated by the fact that the subsamples may actually differ significantly in the overall complex of abilities measured by the test. One needs to develop a method which separates the overall differences from differences attributable to individual questions; in other words, one needs a method which will provide an estimate of the interaction of items with subsamples.

In this study, the two-factor analysis of variance design with repeated measures on one factor is demonstrated to be appropriate to the problem if item difficulties are first subjected to the arcsin transformation. Application is illustrated using SAT candidates from three different subgroups of the total population--Negro candidates from the Southeast, candidates from small town and rural centers in Indiana and Illinois,

and candidates from centers in the Bronx, New York.

CASSERLY, PATRICIA L. College Decisions on Advanced Placement. I. A Follow-up of Advanced Placement Candidates of 1963 (RDR 64-65, No. 15). May 1965.

The results of a mail survey of all colleges and universities that received May, 1963, AP score reports are presented in this, the first of two reports dealing with college decisions on advanced placement. The report is based on tabulations of returns from 90 percent of the colleges questioned and represents over 97 percent of the potential candidates involved.

Previous studies carried out by Educational Testing Service and elsewhere have indicated considerable variability in awards of placement and/or credit in relation to AP examination scores. There were indications that the variability might be related to regional differences or the relative volume of AP candidates at colleges as well as a product of differences across subject-matter fields. Therefore the data were analyzed to test these tentative hypotheses.

Differences across and within individual colleges proved to be the greatest source of variability. In comparison, differences across regions, college groups (as determined by the number of 1963 AP candidates), and subject fields were found to be of minor importance.

Data on awards of placement and credit to students who had not participated in the formal AP program were also collected and compared with data for the formal AP candidates. Although approximately three times as many non-AP students were given some positive college action, our tabulations showed that only 27 percent of this group was given recognition carrying credit, whereas the achievement of 75 percent of AP students who gained any reward for college-level work done in secondary school was recognized by awards of placement and credit or credit alone.

In the appendices of this report are tables that present the relation of AP examination score to various college actions by region, college group, and subject-matter fields.

CASSERLY, PATRICIA L. College Level Examination Program: Its Meaning to Participants. (RDR 72-73, No. 6). April 1963.

It was pointed out earlier in this paper that the obtained sample is not representative of all CLEP takers. The respondent sample contains a disproportionate number of high scores and, thus, of people apt to be well-disposed toward the program. If this is indeed the case, it seems most likely that had an intensive follow-up been conducted and a 100 percent response rate obtained, the percentage of critical responses would have been much greater. Again, inferring from obtained data, they suggest that the most critical philosophical question to which the College Board and ETS must address themselves in the immediate future is whether the CLEP program can or should continue to try to serve everyone with an interest in assessing either his own academic or vocational achievement or facilitating the attainment of a degree.

Perhaps it would be a greater service to more people to redefine the CLEP target population as the truly "nontraditional" student, ascertain what services and materials members of this group especially need, then work out how these special needs can be met and how recognition for achievement meriting credit by colleges or through certification can be assured for these people without further expense and frustration. In addition, it seems that both the future target population and those outside of it could be spared much frustration and disappointment (and the College Board and ETS, therefore, would be spared much ill will and suspicion) if what the program can realistically offer and what it does not or cannot be expected to do are clearly communicated.

However, when considering the above suggestions, no one should lose sight of the fact that for many people the current program and its philosophy do and have provided a much-needed contribution to their personal, academic, and vocational aspirations. The author feels that by more specifically defining its target population--to itself and others--the program can best realize its potential for such people and thereby serve its designated purpose. For it is not at all clear--to say the least--that "nontraditional" students can be well served without attention to their particular needs.

CASSERLY, PATRICIA L. To See Ourselves as Others See Us: An Evaluation of the Advanced Placement Program by Students From 252 Secondary Schools at 20 Colleges (RDR 67-68, No. 5). June 1968.

During 1966-67 the author asked over 400 college students who had taken courses under the College Board's Advanced Placement Program in high school to evaluate how participation in the program had affected their high school and college experiences. The majority of those interviewed were upperclassmen, because it was assumed that they could best evaluate the program in the light of their total educational experience. This is a report on their responses and includes many of their suggestions on improving the AP program at both the high school and the college levels.

CASSERLY, PATRICIA L., AND RICHARD E. PETERSON. College Decisions on Advanced Placement. II. A Report of Interviews Concerned with Advanced Placement Policies and Practices at 63 Colleges and Universities (RDR 64-65, No. 16). June 1965.

Abstract not available.

CASSERLY, PATRICIA L., RICHARD E. PETERSON, AND WILLIAM B. COFFMAN. College Decisions on Advanced Placement. II. An Interview Survey of Advanced Placement Policies at 63 Colleges (RDR 65-66, No. 4). November 1965.

This report, based on interviews with over 500 individuals at 63 colleges and universities, discusses details of institutional policy with respect to the Advanced Placement Program of the College Entrance Examination Board, outlines the mechanics by which institutional policy is implemented, analyzes the problems of communication which in many cases prevent the development and implementation of sound policy, and assesses the overall impact of the Advanced Placement Program on college curricula and on faculty perceptions.

The general picture is one of wide variability in policy, in mechanics, in communication patterns, and in impact. To some extent differences in policies and practices appear consistent with differences in institutional purposes and in characteristics of the student bodies. In significant numbers of cases, however, the observed differences in responses to the Advanced Placement Program reflect assumptions based on misinformation or no information.

A major generalization growing out of the study is that as college faculties become aware of the actual nature of the program, particularly through direct contact with Advanced Placement students, their reservations diminish. There is reason to be optimistic about the future of the program.

CENTRA, JOHN A. Black Students at Predominantly White Colleges: A Research Description (RDR 69-70, No. 2). March 1970.

How do the background characteristics, activities, goals, and perceptions of Black students at predominantly white colleges differ from their white counterparts? In an attempt to answer these questions, research results from several sources were combined in this report. Emphasized in particular were findings from the "Questionnaire on Student and College Characteristics" (QSCC). A sample of 249 Black students at 83 traditionally white institutions had responded to the QSCC in 1968. The similarities and differences between this group of Black students and a matched group of white students are discussed and related to other research evidence.

CENTRA, JOHN A. College Enrollment in the 1980s: Projections and Possibilities (RDR 78-79, No. 1). September 1978.

Growth and size have long been hallmarks of progress in America, and by these standards, higher education has been a resounding success. Enrollments at colleges and universities have climbed to record levels in the past 25 years. In the 1960s alone, undergraduate enrollment doubled to 4 million and total enrollment rose to 8.6 million. A number of conditions have combined to produce this growth. Foremost is a population boom that doubled not only the college-age population between 1953 and 1977 but also the demand for college-trained teachers at all levels. The market for college-educated people was further fueled by

extraordinary expansion in the economy. Pressures for equality of opportunity have also brought unprecedented numbers of Blacks, women, and disadvantaged students to the campuses during the past decade.

Recently, however, much of this has changed or is in the process of changing. The annual double-digit increases in enrollment that were common in the 1960s have shrunk to the more typical 2 to 4 percent annual growth of the 1970s; and much of this growth can be attributed to older students. Most projections for the next 10 or 20 years, as will be discussed later in this report, forecast enrollment declines. Growth in the sectors of the economy that draw most heavily on college graduates has already slowed substantially. And the size of the college-age population--the most important factor affecting future college rates--will shortly begin to decline. By 1990, as Figure 1 illustrates, the 18-to-21-year-old population will return to its 1970 level, just above 4 million. The decrease between 1980 and 1990 will be about 18 percent. The drop during the first half of the coming decade is especially precipitous at 11 percent. By 1985 there will be 1.7 million fewer 18-to-21-year-old population and is expected to drop to about 13 million, representing almost a 25 percent decline from the 1979 peak.

A second important factor in determining enrollment is the degree of attendance or enrollment rate for the college-age cohort and other groups. Will enrollment rates increase enough during the coming years to compensate for the decline in the college-age population? Forecasters differ widely on this question.

The purpose of this report is to analyze the various projections made for the 1980s and their underlying assumptions. The forecasts are essentially based on extrapolations of recent trends for different groups, plus some educated guesses about future conditions. Each of the forecasts is compared and contrasted to determine whether some general agreement exists. Forecasts by region of the country and for different types of institutions are also discussed. Finally, possible new clientele and other ways proposed to increase college enrollments are considered.

CENTRA, JOHN A. The College Environment Revisited: Current Descriptions and a Comparison of Three Methods of Assessment (RDR 70-71, No. 1). August 1970.

The "Questionnaire on Student and College Characteristics" (QSCC), designed to provide information about colleges that would be of interest to prospective students, was administered to upperclass students at over 200 institutions. The 135 items in the instrument elicit student perception and student self-report kinds of information relating to each institution. The purposes of this study were to gain a better understanding of QSCC item relationships, to identify dimensions that differentiate among four-year institutions, and to compare methods of assessing college environments.

This report consists of three sections. In the first section, a factor analysis of the 77 perceptual items in the QSCC is presented and discussed. In the second section, student perception and student self-report data from the QSCC were combined with published, objective information about each institution in order to investigate differences among college environments further. A total of 53 institutional characteristics were factor-analyzed, resulting in six factors: athletic vs. cultural, size with cliquishness, elitism, activism with flexibility, student satisfaction, and social life.

The three methods of assessing the college environment--student perceptions, student self-reports, and objective institutional data--were also compared by use of multimethod factor analysis, a new technique which removes method variance by focusing on correlations between rather than within methods of measurement. These results are presented in the third section of this report.

Several of the appendices in the report include additional analyses of items in the QSCC.

CLEARY, T. ANNE. Test Bias: Validity of the Scholastic Aptitude Test for Negro and White Students in Integrated Colleges (RDR 65-66, No. 18). June 1966.

For this research, a test was said to be biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed,

consistent nonzero errors of prediction are made for members of the subgroup. Samples of Negro and white students from three integrated colleges were studied. In the two eastern colleges, no significant differences in the regression lines were found. In the one college in the southwest, significant differences were found, but it was the Negro scores that were overpredicted. Thus, in one of the three schools, the Scholastic Aptitude Test was found to be slightly biased, but biased in favor of the Negro student.

CLEARY, T. ANNE, AND THOMAS L. HILTON. An Investigation of Item Bias (RDR 65-66, No. 12). April 1966.

For this research, bias was defined as an item x group interaction. Grade 12 students in integrated high schools who had taken two forms of the PSAT were divided into two races and three socioeconomic levels within the races. Four analyses of variance were performed: one for the verbal section and one for the mathematical section of each of the two forms of the PSAT. Because of the large sample sizes used, the tested effects were expected to be and were significant. However, only a minimal percentage of the total variance was contributed by the item x group interactions. On this basis, it was concluded that, if PSAT scores are discriminatory, the discrimination is not largely attributable to particular items, but to the test as a whole.

CLIFF, ROSEMARY, AND NORMAN CLIFF. Formulating Hypotheses: Factor Structure and Validity Data (RDR 63-64, No. 2). September 1963.

It was hypothesized that performance on the Formulating Hypotheses Test is related to the induction, sensitivity to problems, ideational fluency, spontaneous flexibility, and originality factors identified in previous research and to measures of creative achievement in the sciences. A factor analysis of reference tests for the above-named factors failed to identify any but induction. Scores on formulating hypotheses had loadings of borderline significance on fluency and general information factors. No relationship was demonstrated between formulating hypotheses scores and grades of mechanical engineering students for courses stressing creative design.

COFFMAN, WILLIAM E. A Factor Analysis of the Verbal Sections of the Scholastic Aptitude Test (RDR 65-66, No. 17). June 1966.

The purpose of the research reported here was to make a preliminary exploration of the factor structure of the SAT verbal test as a first step in the development of specifications for the test in terms of factor structure. The principal axis factors of a modified tetrachoric covariance matrix with communality estimates based on an interactive procedure were rotated to orthogonal simple structure using the quartimax and the equamax methods. The procedure was repeated twice, once with iterations based on the assumption of 7 factors and once with iterations based on the assumption of 11 factors.

The equamax rotation of the 11-factor solution permits an interpretation in terms of a speed factor, 3 vocabulary factors, 2 reasoning factors, 4 reading skills factors, and a residual factor. The equamax rotation of the 7-factor solution is compatible with the 11-factor solution and indicates that differentiation in the vocabulary, reasoning, and reading areas is a function of item content. The quartimax rotation of the 7-factor solution permits an interpretation in terms of a general verbal aptitude factor accounting for over 72 percent of the common variance, a speed factor accounting for 9 percent of the variance, a small vocabulary factor accounting for 4 percent of the common variance, and 4 residual factors.

COFFMAN, WILLIAM E. On the Validity of Essay Tests of Achievement (RDR 65-66, No. 9). April 1966.

Much has been written regarding the reliability or lack of reliability of essay tests of achievement. In general, the conclusion of investigators is that the typical essay test is limited in that the number of questions that can be answered in a given period of time places a strict limitation on the reliability that can be attained even if methods of framing questions and of grading the responses are highly refined. To achieve high reliability for an essay examination,

it is necessary to make it so long as to be unmanageable in any large-scale evaluation project. Therefore, the practical test specialist will devote his time to the development of more efficient testing procedures and use essay examinations as criteria against which to evaluate them.

But this line of reasoning ignores a fundamental point: how does one achieve the highest validity of measurement? Data obtained in a recent study have raised serious doubts regarding the wisdom of a position that emphasizes reliability at the expense of validity. Single 20-minute essays having reliability coefficients (in the parallel-form sense) of approximately .26 were found to correlate approximately .45 with a criterion measure consisting of four other essays. When reliability was increased to .52 by multiple reading, the correlation with the criterion increased to .66. This finding, while contrasting markedly with the findings from typical validity studies, is shown to be consistent with test theory as formulated in the correction for attenuation.

It is argued that whenever the items in an achievement test may be considered a representative sample of the criterion performance, the validity of the test may be estimated by applying the correction for attenuation. The implications are set forth using data from the empirical study to establish anchor points and developing estimates of the reading reliability, score reliability, and validity of essay scores based on different combinations of topics and readers.

COFFMAN, WILLIAM E. A Report on the Pretesting for the Scholastic Aptitude Test in Spanish-Form APR (RDR 63-64, No. 12). April 1964.

Abstract not available.

COFFMAN, WILLIAM E. The Scholastic Aptitude Test of 1963-64 (RDR 63-64, No. 15). January 1964.

Abstract not available.

COFFMAN, WILLIAM E., AND DANA G. KURFMAN. Single Score Versus Multiple Score Reading of the American History Advanced Placement Examination (RDR 65-66, No. 14). May 1966.

Four readers assigned to read a single essay question of the Advanced Placement Examination in American History in 1962 participated in two experiments designed to assess the differential effects of reading by an analytical method by which three different scores were assigned to a paper and by a holistic method by which a single score was assigned. Because of limited time and the small number of readers, the two experiments were conducted in an interlocking design that permitted the findings of one experiment to sharpen the interpretation of the effects observed in the other.

Analyses of variance procedures revealed significant effects attributable to differences in standards among the four readers, to differences in standards from day to day, and to fluctuations in standards of individual readers from time to time during the reading period. There was no definite evidence of differences in reading reliability from method to method.

COFFMAN, WILLIAM E., AND MARGARET H. MAHONEY. The Relationship Between Scores on the Scholastic Aptitude Test and Certain Post-College Activities of Phi Beta Kappa Members of the Yale Classes of 1931-1950 (RDR 67-68, No. 1). July 1967.

The study attempted to describe relationships between post-college activities of members of Phi Beta Kappa at Yale University and their SAT scores. Three types of data were tabulated by levels of SAT scores: occupation followed by the individuals, level of achievement reached by each individual in most of the occupational groups, and frequencies of graduate or professional degrees. Two additional relationships were examined. The frequencies of graduate or professional degrees combined with frequencies of occupational listing computed for each level of SAT scores and the number of individuals listed in Who's Who in America, American Men of Science, or the Dictionary of American Scholars comparing their mean SAT verbal scores with the mean SAT-V scores of all Phi Beta Kappans in the corresponding college class.

The limitations of the study precluded generalization from the available data to relationships which might be expected by the currently obtained SAT scores and post-college activities over the next 20 years. However, those data did raise questions and pointed up the need for long-range follow-up studies of more representative samples.

COFFMAN, WILLIAM E., AND MARY ELLEN NEUN. Effects of an Accelerated Reading Course on SAT-V Scores (RDR 65-66, No. 8). April 1966.

Three groups of university freshmen took the SAT-V before and after a course in accelerated reading. The mean posttest scores for two groups increased slightly less than 10 points, while the mean posttest score for the third group decreased significantly. It is concluded that for the type of students used in this study, a course emphasizing speed and accuracy of reading is not of value in increasing scores on the verbal part of the Scholastic Aptitude Test.

DAVIS, JUNIUS A. Faculty Perceptions of Students, I. The Development of the Student Rating Form (RDR 63-64, No. 7). February 1964.

The steps leading to the development of a Student Rating Form, an instrument for use by faculty in systematically describing college students, are outlined in this first of a series of reports concerned with faculty characterizations of students. The treatment is primarily historical, providing a basis for subsequent technical reports of analyses of faculty definition of desirable student traits. Data presented include a thematic analysis of free verbal descriptions by faculty of students, and the results of a factor analysis (suggesting 16 factors) of variables contributed by an early version of the Student Rating Form.

DAVIS, JUNIUS A. Faculty Perceptions of Students, II. Faculty Definition of Desirable Student Traits (RDR 63-64, No. 8). March 1964.

The primary purpose of this study was to determine student traits that faculty associate with desirability, separate and apart from those reflected by the traditional academic achievement indices.

Ratings of students on 80 variables, including a student desirability variable, were obtained, together with high school and college grade averages and SAT scores. The resulting 84 x 84 correlation matrix was factored by the diagonal method, using precise communality estimates from a separate factor analysis. In this case, the procedure permits partialing out the variance in desirability and the other variables which is attributable to academic performance, and definition of residual desirability in terms of relationship of the residual of the other variables to desirability apart from academic performance. The diagonal factoring method was also used to examine the content, including academic performance, of general desirability, and to define desirability separate and apart from SAT and academic performance.

Ratings of intellectual ability and values, motivation, and creativity, as well as actual grade point average, were found to be related to general desirability ratings. Desirability apart from grades, however, appears to consist of such traits as likableness, ethicality, open-mindedness, altruism, maturity, and self-insight, although ratings of intellectual ability and values have components related to grades and to desirability apart from grades. Ability as measured by the SAT, though reasonably related to performance, appears to have negative relationship to desirability apart from grades.

DAVIS, JUNIUS A. Faculty Perceptions of Students, III. Structure of Faculty Characterizations (RDR 63-64, No. 9). April 1964.

The major purpose of this study was to determine the structure of characterizations of students by faculty members. What dimensions (or trait patterns) are of concern, or are observable, and appear meaningful?

Ratings of students by faculty members on 80 rating scale variables, together with SAT scores and high school and freshman average grades, were obtained; intercorrelations were computed; and the resultant matrix factored and rotated to the normal equamax criterion.

Eighteen factors were found, 16 of which were drawn primarily from the rating

scale variables. The 18 factors accounted for about 86 percent of the correlation among the original variables. These factors reveal a relatively complex or multidimensional judgmental framework, with factors devoted to various aspects of intellectual qualities, functional orientation to work, character traits, interpersonal relationships, and intrapersonal adjustment.

DAVIS, JUNIUS A. Faculty Perceptions of Students, IV. Desirability and Perception of Academic Performance (RDR 63-64, No. 10). March 1964.

A previous study in this series was devoted to the definition, using the diagonal method of factoring, of student desirability separate and apart from academic performance. Variables employed consisted of ratings of students by faculty on 80 traits, as well as high school and college grade averages and SAT scores.

In the present study, the same data and similar methods were employed to determine the nature of desirability separate and apart from the rater's perception of the student's academic performance (rather than his actual performance).

The results tend to show an absorption in the faculty perception of performance of much of the variance previously associated with dependability, intellectual ability and values, motivation, self-sufficiency/creativity, and maturity in defining desirability apart from academic performance. The results tend, however, to verify or accentuate the contribution of other traits, notably ethicality, likableness, open-mindedness, and altruism. It does appear that desirability is highly related to perception of academic performance, and that other areas normally expected to make an honest contribution to performance (e.g., dependability, intellectual ability and values, motivation, and self-sufficiency/creativity) are perceived by the faculty do so. Finally, there is remaining reliat variance in SAT and actual grades, as well as in such previously established areas as conformity and extraversion, which is related neither to perception of performance nor desirability.

DAVIS, JUNIUS A. Faculty Perceptions of Students, V. A Second-Order Structure for Faculty Characterizations (RDR 64-65, No. 14). May 1965.

A matrix of 80 different trait ratings of college students by teachers, plus high school and college grades and the students' SAT scores, was factored, utilizing the Promax method of Hendrickson and White. This yielded 18 factors. From intercorrelations among the factor axes, a principal axes procedure was used for (second-order) factor extraction, and four factors were defined by rotation to the normal Equamax criterion. These factors were tentatively named (I) Interpersonal Style, (II) Ethical Orientation, (III) Intellectual Involvement, and (IV) Academic Performance Style.

DAVIS, JUNIUS A. Faculty Perceptions of Students, VI. Characteristics of Students for Whom There Is Faculty Agreement on Desirability (RDR 65-66, No. 15). June 1966.

This final report in the Faculty Perceptions of Students series has concerned, in the main, examination of students defined by a consensus of faculty as desirable. A variety of measures were tested for relationship against criteria drawn from frequency of being known and nomination by faculty to several laudatory categories. The major component of general acclaim was revealed to be academic performance, although technical difficulties (size of sample, limited reliability of other criterion measures) may have dampened other potential findings.

EKSTROM, RUTH B. Colleges' Use and Evaluation of the CEEB Writing Sample (RDR 63-64, No. 5). January 1964.

A sample of 23 colleges that have required the CEEB Writing Sample was visited. Of the 20 colleges using the Writing Sample for admissions, 11 read for a general impression and 9 assigned grades. The qualities looked for most often were: (1) organization, (2) indication of ability to think, (3) spelling, and (4) grammar. In responding to a check list, clarity of expression, organization, correct grammar, logic of reasoning, and evidence of mature thought were selected as

most important. Experimental studies carried out by some of the colleges showed that the Writing Sample had low reliability and validity as an admissions criterion. Colleges using the Writing Sample for placement in English courses usually use it for selecting advanced students or for selecting students needing remedial work in composition. If a centrally graded theme were to be offered as a part of the English Composition Test, the Writing Sample could probably be abandoned without too much discontent among CEEB member colleges. Neither the enthusiasm for nor the effectiveness of the Writing Sample is very great.

EVANS, FRANKLIN R. A Study of the Relationships Among Speed and Power, Aptitude Test Scores, and Ethnic Identity (RDR 80-81, No. 2). October 1980.

This is the fourth report in a series of research studies to determine the effects of altering time-per-item as it relates to the test candidate's ethnic identity, residence (rural or urban), and sex. Four specially designed SAT-like tests were administered to volunteer samples of Black, Chicano, and white college-bound juniors in twelve high schools in urban and rural settings. Within each high school, the volunteers were randomly assigned to one of three time-limit conditions: speeded (20 minutes per section), normal (30 minutes per section), and unspeeded (40 minutes per section).

Results of several independent analyses of variance (ANOVA's) did not reveal any differential performance related to residency or sex. Significant effects were detected for test-time limits and ethnic identity. No significant two-way interactions involving any of the four factors (time, ethnic identity, residence, or sex) were detected.

Two significant findings were observed in the item analysis of the selected sections (verbal, math). There is some evidence that minority volunteers were employing less than optimal test-taking strategies. The average section scores of the volunteers "still working" near the end of the test section were often lower than the average sections scores for all candidates. Second, when alternate form reliability estimates were compared for minorities, the estimates of reliabilities for wrong scores were higher than those for rights or formula scores. Possible explanations of these phenomena are that the tests were too difficult for the volunteers or that the volunteers who persisted were being attracted to incorrect options by way of misinformation.

FLAUGHER, RONALD L. Project Access Research Report No. 2, Patterns of Test Performance by High School Students of Four Ethnic Identities (RDR 70-71, No. 9). May 1971.

Among the four participating ethnic groups in Project Access for Los Angeles (Black, white, Mexican-American, and Oriental), the patterns of test performance that emerged were surprisingly similar to those from two previous studies of first-grade children, even though Access involved eleventh graders. Partially replicated was the finding that overall level of performance varied with social class, while the patterns remained rather constant within ethnic group, regardless of social class.

This new evidence adds to the strength of the argument that our educational systems should be attending to, and utilizing, the variety of talents and aptitudes presented by the students, rather than attempting to treat everyone as a member of the majority group.

FLAUGHER, RONALD L. Project Access Research Report No. 3, Minority Versus Majority Group Performance on an Aptitude Test Battery (RDR 71-72, No. 1). August 1971.

Black students from two cities and Mexican-American students, both male and female, showed small but consistent tendencies to perform better, relative to white groups, on three nontraditional measures: tests of inductive reasoning, spatial scanning, and associative memory. These measures showed somewhat less discrepancy between the groups than did tests of the more traditional verbal and mathematical aptitudes.

The results seemed compatible with the suggestions of the Commission on Tests to expand the number of measures included in traditional testing programs. Other measures of these same aptitudes, and their validities, remain to be explored in this context.

FLAUGHER, RONALD L., AND LEWIS W. PIKE. Reactions to a Very Difficult Test by an Inner-City High School Population: A Test and Item Analysis (RDR 69-70, No. 4). June 1970.

A frequent complaint of minority and poverty-level groups in our society is that of the "irrelevance" of many of the institutional practices which they encounter in their efforts to upgrade the quality of their existence. Educational institutions certainly have not been exempt from this accusation, and testing practices, in particular, have been the topic of some criticism.

It was the purpose of this paper to examine just one aspect of this problem of appropriateness of testing practices. Through test and item analysis procedures, it examines the performance of a low-achieving, inner-city high school population on a machine-scored, multiple-choice examination which had been designed for and directed toward a majority-group, basically middle-class population.

FLAUGHER, RONALD L., AND DONALD A. ROCK. Project Access Research Report No. 5, Patterns of Ability Factors Among Four Ethnic Groups (RDR 71-72, No. 9). June 1972.

The similarity among the factor loading patterns would indicate that essentially the same patterns of abilities are being demonstrated regardless of the ethnic identity of the examinee. The inclusion of a wider variety of measures, or a different selection of ethnic groups, might yield different results. However, this study seems to indicate that, at this age level, the factorial structure elicited by objective examinations does not differ significantly among the more predominant ethnic groups of the country. Until evidence to the contrary can be discovered, the assumption of similar structures seems justified.

FLAUGHER, RONALD L., AND DONALD A. ROCK. The Wide Range Validity of Certain New Aptitude Tests (RDR 66-67, No. 8). December 1966.

A study by French showed that the addition of certain new types of test material to the College Board Scholastic Aptitude Test (SAT) would result in significant improvement of the validity of the SAT for the prediction of college performance by students of high ability. The present study examined the validity of these same types of test materials over the wider range of ability that characterizes college freshmen in College Board member colleges.

FREMER, JOHN, WILLIAM E. COFFMAN, AND PHILLIP H. TAYLOR. The College Board Scholastic Aptitude Test as a Predictor of Academic Achievement in Secondary Schools in England (RDR 66-67, No. 9). May 1967.

A major difficulty in constructing aptitude tests is that, for certain groups of candidates, irrelevant content factors may result in systematic score differences that are unrelated to performance on criterion measures. This problem becomes even more prominent when tests developed for use with one group are considered for use with groups that have significantly different characteristics. In the course of increasing internationalization of education, tests developed in the United States have been used in other countries. This practice raises questions as to the extent to which cultural differences will result in decreased efficiency of the transplanted tests. This study used several techniques to assess the validity of the College Entrance Examination Board's Scholastic Aptitude Test (SAT) for students in selective secondary schools in England. The questions to be answered were: What kinds of distributions of scores are obtained? How reliable is the SAT for this population? How well do the individual items work? How well do the test scores correlate with academic criteria?

A form of the SAT was administered to students in eight secondary schools in England. Data from a representative sample of College Entrance Examination Board candidates in the United States were also available. The British students performed very well on the SAT. The level of reliability, item biserial correlations, and correlations with academic criteria were comparable to those found in studies in the United States. Some verbal items, however, had substantially different levels of difficulty for the British group.

The results suggest that United States tests can work well in other countries but that certain items in any test might have to

be changed to take into account cultural differences.

FRENCH, JOHN. The Revision of the Interest Index (RDR 64-65, No. 5). October 1964.

Abstract not available.

GITLITZ, ALFRED H., AND NADEEN L. KAUFMAN. Project Access Research Report No. 4, Influence of Race, Sex, and City on Inductive Reasoning Items (RDR 71-72, No. 8). June 1972.

Letter Groups is one of nine subtests in the Project Access battery. It is composed of 25 items, which the examiners assigned to different categories according to the particular type of cognitive task the item's solution demanded. Analyses then proceeded with comparisons of item difficulty levels within race, city, sex, and race and city groupings. This was done to determine whether certain item categories were differentially more difficult for some groups than were other categories.

Results indicated that the majority group (whites) fared better on Letter Groups than did minority groups (Mexican Americans or Blacks). However, on one category of items there was no difference between majority and minority performance, whereas on one category of items there were greater discrepancies in performance between these groups. The authors raised the question that the distribution of items on a test by their different categories (and therefore different cognitive demands) could strongly influence how groups perform.

GODSHALK, FRED I., FRANCES SWINEFORD, AND WILLIAM E. COFFMAN. Studies of the Question Types in the CEEB English Composition Test as Predictors of an Essay Criterion (RDR 64-65, No. 13). April 1965.

This series of studies was initiated originally to assess the relative validities of objective and semi-objective types of questions developed for the English Composition Test since 1947. Of particular interest was the Interlinear Exercise, a type highly valued by the Committee of Examiners but widely criticized by teachers in schools and colleges. During the course of the first study, evidence was assembled pointing to unique validity for a 20-minute essay, and later stages of the investigation focused on further study of the essay as a predictor.

These studies differ from previous studies in that the criterion is a series of essays, read by a number of different readers, rather than the typical criterion of English grades or ratings by classroom teachers. For the earlier phases of the study, the criterion was the total of 25 judgments, five on each of five different essays; in later phases, the criterion was the total of 20 judgments one of the essays having been removed and used as a predictor.

The method of reading essays may be described as holistic rather than atomistic or analytical. Each reader read each essay rapidly and assigned a single rating. The original ratings were on a three-point scale; later a four-point scale was used. Readers made between 30 and 40 judgments per hour.

The subjects were 646 eleventh- and twelfth-grade students in 24 schools, 533 of whom had taken the PSAT and 138 of whom took the SAT prior to the analysis of the data. The group was slightly less able than the typical group of SAT candidates but similar to it in variability. Each subject wrote the five essays, worked six 20-minute objective subtests (paragraph organization, sentence correction, usage, prose groups, error recognition, and construction shift), and completed two 30-minute Interlinear Exercises during a three-week period in October and November, 1961.

A detailed analysis of the data supports the following conclusions:

1. The sum of 25 readings of five essays is sufficiently reliable to provide a stable criterion measure. Reading reliability is .921; score reliability is .841.

2. The paragraph organization type of question has low validity. A 20-minute subtest correlates with the criterion less than .50 and adds nothing significant when combined with other types of questions.

3. With the exception of paragraph organization, the other objective and semi-objective types are effective as predictors of the essay criterion. The validity coefficients range from .568 to .709 for the total sample of 646 cases, and combinations of three types in one-hour tests produce correlations ranging from .716 to .775. Combinations which include an Interlinear Exercise tend to produce the higher multiple

correlations.

4. When PSAT is added to the composite, the multiple correlations tend to increase; however, first-order correlations for the PSAT with the criterion tend to be lower than those for some of the 20-minute subtests.

5. There appears to be no significant interaction between level of ability of subjects and the relative validities of the several types of questions; in other words, questions seem to work with equal effectiveness for able and less able students.

6. When one of the 20-minute essays is treated as a predictor and correlated with the sum of ratings on the other four essays, the correlations are consistently higher than the estimates of score reliability for the predictor essay. When 20-minute essay scores are based on as many as three readings, the validity coefficients are of the same magnitude as those for the machine-scorable subtests.

7. The 20-minute essay contributes uniquely to the prediction of the essay criterion. Validity coefficients for combinations which include an essay score are consistently higher than combinations which do not include the essay.

8. The 20-minute essay is no less valid when scores are obtained under field conditions simulating large-scale reading.

9. It is estimated that predictor essays can be processed at a cost approximately the same as that for reading Interlinear Exercises.

GRANDY, JERILEE E. An Investigation into the Construct Validity of Three Career Skills Exercises (RDR 79-80, No. 1). December 1979.

Abstract not available.

HALPERN, GERALD. Scale Properties of the Interest Index (RDR 65-66, No. 5). November 1965.

Extensive research by John French on the Interest Index produced a 12-scale interest test with 16 items per scale. The present study investigated certain scale properties of the Interest Index. Internal consistency coefficients were found to cluster about .91, item-scale assignment was verified for 99 percent of the items, scale test-retest correlations (three-week interval) were found to cluster about .86, and the individual scales were found to be remarkably independent of one another. It was concluded that, on all these measures, the Interest Index performs at least as well as any comparable test and, in addition, has three advantages over most. It is a very efficient instrument (in terms of number of items per scale), it utilizes subject-matter experiences familiar to most students, and it measures interest in a wide variety of subject-matter course areas. Partial confirmation of the construct validity of this test is provided, and it is recommended that (a) appropriate norms be developed and (b) further validation studies be conducted.

HALPERN, GERALD, AND MASU SASAJIMA. Effectiveness of PSAT and SAT Booklets, I. Bulletin for Students, 1964 Preliminary Scholastic Aptitude Test (RDR 64-65, No. 19). June 1965.

A questionnaire was developed and distributed to over 600 students in 19 different schools. The purpose of the questionnaire was to evaluate the effectiveness of the Bulletin for Students, 1964 Preliminary Scholastic Aptitude Test. Additional information was obtained on school practices in handling the Bulletin.

The analyses of student responses indicate that the facts and concepts presented in the Bulletin are comprehensible, in general, to high school juniors. This is less true for students with relatively low PSAT scores. The most difficult concept for the students is the correction-for-guessing formula applied when scoring the PSAT. In general, students view the Bulletin very favorably.

HALPERN, GERALD, AND MASU SASAJIMA. Effectiveness of PSAT and SAT Booklets, II. Your College Board Scores: Preliminary Scholastic Aptitude Test (RDR 65-66, No. 3). October 1965.

Students in 19 different schools were asked to answer a number of questions in order to evaluate the comprehensibility of the booklet entitled Your College Board Scores: Preliminary Scholastic Aptitude Test. Five

hundred sixty-five students, selected at random within these schools, completed the questionnaire.

Information is provided on students' comprehension of the booklet, including their ability to convert test scores to percentiles and to predict SAT scores from a knowledge of PSAT scores. The perceived implication, by students, of the use of the PSAT for college admissions and for the prediction of college success is discussed.

The booklet, in general, was comprehensible to the majority of students. However, some of the facts and concepts presented (e.g., prediction of SAT scores, use of interpolation, interpretation of a percentile) did prove difficult for many of the students. Comprehension was found to be positively related to scholastic aptitude.

HALPERN, GERALD, AND MASU SASAJIMA. Effectiveness of PSAT and SAT Booklets, III. A Description of the College Board Scholastic Aptitude Test (RDR 65-66, No. 11). April 1966.

A questionnaire was developed and distributed to over 500 students in 19 different schools. The purpose of the questionnaire was to evaluate the effectiveness of A Description of the College Board Scholastic Aptitude Test. Additional information was obtained on school practices in handling this booklet.

It was concluded that the booklet has been found to be both understandable and attractive to most of the students who participated in this study. These 17 questions that were scored were, on the average, answered correctly by about 70 percent of the students. Booklet comprehension was associated with scholastic aptitude. Several of the facts and concepts presented (e.g., when to guess, order of difficulty of the items) were misunderstood by the students and should be revised in later editions of the booklet.

HALPERN, GERALD, AND MASU SASAJIMA. Effectiveness of PSAT and SAT Booklets, IV. Your College Board Scores: Scholastic Aptitude Test, Achievement Tests (RDR 65-66, No. 16). June 1966.

A questionnaire was developed and administered to about 450 twelfth-grade students as a means of evaluating the booklet, Your College Board Scores: Scholastic Aptitude Test, Achievement Tests. In general, the booklet was found to communicate effectively. Exceptions to this general statement were difficulties in locating appropriate tables and a superficial understanding of the meaning of test scores and percentiles. Comprehension of the booklet was moderately related to scholastic aptitude scores.

HARRISON, SYLVIA E. A Summary of Current Social Studies Curriculum Projects (RDR 64-65, No. 8). November 1964.

This is a summary of projects that have national implications for the social studies curriculum. No specific mention is made of projects concerned with curriculum development at the college level or of the considerable experimentation and revision that was taking place at the local level.

Material received from directors of projects was organized as follows: a statement of objectives followed where appropriate a brief description of the procedures undertaken by the project to reach its objectives, the project's activities to date, the course content involved, citation of the organizations for each project, and duration of the project.

HARTNETT, RODNEY T. Learning Style Preferences Among College Students (RDR 72-73, No. 5). February 1973.

By examining data gathered from college students on three different occasions (at the time they entered college, at the end of their freshman year, and at the end of their sophomore year), this paper describes trends in student learning-style preferences and identifies various characteristics of students associated with these preferences.

HAVEN, ELIZABETH W., AND ROBERT E. SMITH. Financial Aid to College Students, 1963-64 (RDR 64-65, No. 11). March 1965.

This survey examines institutionally administered financial aid awarded to full-time undergraduates in 1,221 colleges for the academic year 1963-64. The authors establish that the total sum included about $120 million in scholarships, $85 million in loans, and $45 million in jobs. They also describe the various forms of distribution of aid by geographic region, type of institution, type of students, amount and type of aid received, etc. This important study now serves primarily as a benchmark because the conditions of student financial aid changed so radically in the 1960s.

HEIDGERD, LLOYD H. A Survey of College Physics Departments Regarding Advanced Placement Practices (RDR 64-65, No. 7). October 1964.

A survey was made of nine colleges that enrolled half the Advanced Placement Physics candidates who entered colleges in the fall of 1960 and the fall of 1961. Of the 558 students included, 335 had Advanced Placement Physics Examination grades of good, honors, or high honors (scores of 3 or better). However, the data were indeterminate concerning 54 of these (these are students who took no physics according to the data received and who were not awarded advanced standing--it is not known whether they were awarded advanced standing at a later date). Of the remaining 281: 56 (20 percent) were given one semester advanced standing; 32 (11 percent) were given two or more semesters of advanced standing; 22 (8 percent) were given advanced standing of an indeterminate amount; and 171 (61 percent) were given no advanced standing, but 58 of them (34 percent) were allowed to enter an honors course along with other students who gave evidence of having outstanding ability.

Considering the results of the survey as a whole, it is evident that during the years covered, Advanced Placement Physics did not result in as much advanced standing as might have been expected. Despite good scores in the Advanced Placement Physics Examination, many students received no advanced standing at the time of their enrollment in college. It must be recalled that the survey dealt only with those colleges that received the largest numbers of Advanced Placement students, and it is possible, therefore, that more recognition may have been given to Advanced Placement students by the colleges that received smaller numbers of candidates. That this may be the case is suggested by the fact that the smaller colleges covered by the survey did appear to be more liberal in awarding standing than did the larger colleges. This trend is an interesting contrast to the general experience of the Advanced Placement Program that the colleges which receive the most candidates tend to be the most liberal in awarding advanced standing.

The data in this survey also indicate rather wide variation among the colleges in advanced placement practices. A candidate with an Advanced Placement Physics score of 4 or 5 was almost always given advanced standing if he enrolled in some of the colleges surveyed, but this was clearly not the case if he enrolled in others. This suggests that an Advanced Placement student who is seriously desirous of obtaining advanced standing for his work in an Advanced Placement Physics course would be well advised to ascertain the policies and practices of the colleges he is considering entering, since some colleges seemed rather reluctant to grant advanced standing to more than a very few persons.

It also appears from the data of the survey that a student is more likely to obtain advanced standing if he can offer a good score in Advanced Placement Mathematics in addition to Physics. In fact, in two colleges, high scores in both subjects were necessary for advanced standing.

HILTON, T. L., AND A. E. MYERS. Growth Study, II. Personal Background, Experience, and School Achievement: An Investigation of the Contribution of Questionnaire Data to Academic Prediction (RDR 65-66, No. 7). Februarv 1966.

Numerous studies have demonstrated that school achievement can be predicted from biographical data. This study was designed to determine whether biographical data contribute anything to the prediction beyond that provided by a battery of ability and achievement tests.

The data were obtained from a continuing longitudinal study of academic growth

initiated at ETS in 1960. As part of this eight-year study, questionnaire and test data were obtained from approximately 5,000 eleventh graders in 17 United States public school systems. A year and a half later, test scores and academic standing were obtained for these students as seniors.

The present study includes only 1,206 boys who were graduated from academic programs and for whom complete data were available. The objective test predictors were scores from the School and College Ability Test (SCAT), the Sequential Test of Educational Progress (STEP), and a specially designed Test of General Information (TGI).

The questionnaire data concerned the students' experiences during the previous year. From a unit weighting subsets of the 169 questionnaire items, scales were created to measure variables such as Amount of Low Level TV Watching and Level of Occupational Plans.

Two types of criteria were examined: (1) scores from the Preliminary Scholastic Aptitude Test and the College Board English Composition and American History Tests and (2) each student's rank in his graduating class.

Multiple regression analysis indicated that biographical information, used as the sole source of predictors, predicted both the objective test criteria and rank in class moderately well. When used in conjunction with objective test scores, however, the test scores accounted for practically all the variance in the criteria. Further, factor analysis indicated there were no factors that were common to the biographical variables and the criteria but not common to the objective test predictors and the criteria. Alternative interpretations and implications are discussed.

HORCH, DWIGHT H., AND AMIEL T. SHARON. Estimating Parents' Contribution to College Costs: The Accuracy of Three Measures of Succeeding Year Family Net Income (RDR 70-71, No. 8). May 1971.

The accuracy of each of three measures of future income was investigated for a sample of families who filed Parents' Confidential Statements with the College Scholarship Service in two consecutive years. The three income measures investigated were: (1) parents' reports of current (1968) income, (2) parents' estimates of 1969 income, and (3) predictions of future year (1969) income made by application of least squares regression parameters to parents' estimate of income. Two types of accuracy were considered. The first was the accuracy of prediction assessed by the correlation between the predictors (current or estimated income) and the criterion (actual 1969 income). This analysis indicated that the best single predictor of actual income was estimated income, with a validity coefficient of .86, and was followed closely by current income, with a validity coefficient of .83. The accuracy of each income measure was also investigated by assessing the median discrepancies between parents' contribution associated with each income measure and the contribution associated with actual 1969 income. Current and estimated income result with about equal frequency in substantial (greater than $100) undercalculations of parents' contribution; but use of current rather than estimated income would result in substantial (greater than $100) overcalculations for approximately 31,000 additional CSS families. In addition, needs analysis procedures that rely on either current or estimated income place low-income families at a disadvantage by overstating their future-year income. This study suggests least squares regression as one possible approach to remove this inequity.

HOWELL, JOHN J. College Board Scores of Candidates of Non-English-Language Background Tested at Foreign Centers (RDR 64-65, No. 2). August 1964.

A study was made of the College Board scores of 2,821 candidates of non-English-language background who were tested at foreign centers in December, 1962, or in January, March, or May, 1963. The purposes of this study were to provide normative information and comparisons that might facilitate the interpretation of College Board scores of foreign candidates of non-English-language background, and to assess the language handicap of specified groups on the various tests and explore the implications of these results.

In general, the foreign candidates of non-English-language background who take College Board examinations do well on them despite the language barrier. They have more difficulty with the Scholastic Aptitude verbal and English Composition Tests

than with the Scholastic Aptitude mathematical or the science Achievement Tests. The foreign language Achievement Tests seem inappropriate for a large proportion of those candidates of non-English-language background who actually take them.

In the population studied, male candidates scored consistently higher than female candidates on the two mathematics tests for which sufficient data for comparison were available. On the Scholastic Aptitude verbal test there was little difference in scores between the sexes. Candidates later admitted to college in the United States or Canada tended to be those who had scored well on the tests, but the mean scores of the entire groups tested in grades 11 and 12 were, in most instances, not far below the mean scores of the candidates later admitted to college. Self-ratings by the candidates on knowledge of English were found to be strongly related to scores on the Scholastic Aptitude verbal and English Composition Tests but only slightly related to scores on the Scholastic Aptitude mathematical test. "Language handicap" is defined, the problem of measuring it is discussed, and recommendations for further research are provided.

HOWELL, JOHN J. Language Background of Candidates for College Board Examinations Tested at Foreign Centers (RDR 63-64, No. 1). October 1963.

College Board candidates tested at all foreign centers and at centers in United States overseas territories were asked to fill out questionnaires on language background at each of the four major test administrations in 1962-63. The returns cover 93 percent of the SAT candidates and 70 percent of the Achievement Test candidates. English is the native language of 8,355 of the candidates surveyed (73 percent of the total of 11,467); of these, 3,054 are citizens of the United States, a large majority of the remainder being citizens of Canada.

Of the 3,112 candidates of non-English-language background, 876 reported Spanish as their native language, 523 Chinese, 492 "other European language," 264 French, 120 Arabic, 115 German, and smaller numbers various other specific languages or language categories. Nearly two-thirds know English definitely less well than their native languages according to their own report. Country-by-country tabulations are provided. The implications of the data for the interpretation of College Board scores of foreign candidates for admission to United States colleges and universities are discussed.

JACKSON, REX, AND WILLIAM B. SCHRADER. Verbal and Mathematical Ability of High School Juniors in 1974: A Norms Study of PSAT/NMSQT (RDR 76-77, No. 2). September 1976.

This study was undertaken to obtain current information on the performance of a national sample of all high school juniors on the Preliminary Scholastic Aptitude Test/National Merit Scholarship Qualifying Test (PSAT/NMSQT). The distributions of scores for such a sample are useful in score interpretation because they make it possible to find a student's relative standing in a suitable reference group. Beginning in 1960, results of similar surveys have been used to prepare national norms for the PSAT and SAT. Because the students taking these tests are self-selected with respect to ability, national norms based on all students provide a useful supplement to descriptive statistics based on test candidate groups. The steady decline of SAT scores beginning in 1963-64 suggested the need for new data to replace those obtained from a similar study conducted in 1966.

The sample design for this study was similar to that used in the surveys conducted in 1960 and 1966. Provision was made, however, to replace nonparticipating schools in the initial sample with similar schools. Substantial attention was given to evaluating the representativeness of the sample by comparing participating with nonparticipating schools on the basis of geographical region, control, and percentage of minority students, and by comparing student data for the norms sample with data from the National Longitudinal Study.

The results of this study were compared with those of the 1960 and 1966 norms studies. These comparisons, although not definitive, should be considered in assessing whether a shift in the ability level of high school students generally can help to account for the SAT score decline.

JACOBS, PAUL I. Effects of Coaching on the College Board English Composition Test (RDR 63-64, No. 13). May 1964.

Coaching may increase students' scores on an Achievement Test without increasing, or increasing temporarily, skills in the area being tested. This study deals with the immediate and longer term effects of coaching on the CEEB English Composition Test.

Students at each of six cooperating high schools were randomly assigned to coached and non-coached conditions. Following the coaching sessions, all students took the English Composition Test.

The immediate effects varied from essentially no mean differences in test scores between coached and non-coached students at one school to a mean difference of 73 points favoring the coached students at another school. In general the coaching increment was not confined to the specific item types coached for. The coached groups achieved their superiority by getting few[er] items wrong.

When coached and non-coached students were retested ten months later, it was found that the coaching increment had disappeared. Estimates of growth and practice effects from test-retest data of the non-coached students varied from a mean of 15 points at one school to a mean of 94 points at another.

It was concluded that the candidate who receives coaching may indeed have an advantage over the candidate who does not. But the variability from school to school both of coaching effects and of score gains due to "growth" indicates that the specifics of the teaching situation are quite important.

JONES, CHANCEY O., AND S. IRENE WILLIAMS. College Mathematics Courses Through the Calculus: A Survey (RDR 74-75, No. 1). 1974.

At the request of the College Entrance Examination Board, members of the Test Development Division of Educational Testing Service developed a questionnaire to ascertain the types of mathematics courses through the calculus that are being offered at institutions of higher learning and to ascertain, to a lesser extent, the kinds of mathematics testing instruments that would be useful to the colleges and universities.

The questionnaire was developed during the spring of 1973 and was sent in September, 1973, to a sample of 455 mathematics department chairmen at a variety of universities and two- and four-year colleges.

The data gathered from the questionnaires returned by 252 of the 455 institutions were analyzed and collected into a number of tables which make up the basis of this report. The results of this survey will be made available to a committee of mathematicians who will be convened for the purpose of recommending to the College Board a series of mathematics examinations that will best serve most colleges and universities in awarding credit and placement in mathematics to deserving students.

KARAS, SHAWKY F. Change in Scholastic Aptitude Test Scores of Candidates Tested in the Junior and Senior Years 1961-1964 (RDR 64-65, No. 1). July 1964.

The data collected for the 12 studies summarized in this report indicate that the six March or May to January and the March, 1960, to January, 1961, verbal score gains are lower than any average gains observed previously for studies of comparable time spans. But from the fluctuation of these gains, no consistent trend could be observed. This is probably due to the effect of sampling errors inherent in the procedures used to place successive forms of the SAT on the CEEB scale. For the other six studies, in general, the average score changes are similar to those obtained in previous studies with the exception of the May to December, 1963, average score change.

Neither a trend toward an increase nor a trend toward a decrease in average mathematical score change can be observed in these studies. In general the average score changes are similar to those obtained in previous studies with the exception of the May to December, 1963, average score change. There is no evidence from these studies or from other checks on the scaling of the various SAT forms to indicate that there is any substantial error in the scaling of the current forms of the tests involved in these studies.

In 1963-64, for the first time, the score change data were analyzed separately for boys and girls as well as for total groups. For each of the four 1963-64 studies, broken down by sex, mathematical score average changes are higher for males than females. Neither a trend toward a

higher nor a trend toward a lower average verbal score change for males than females can be observed in these four studies. The average gains on the mathematical score are consistently greater than those of the verbal score for males. But for females, the average gains on the verbal scale are greater than those of the mathematical scale for all the 1963-64 studies except the May, 1963, to January. 1964, study, where they are equal.

KARAS, SHAWKY F., AND LORNE M. KENDALL. Development of a Reduced Set of Composite Equations for Three Predictors (RDR 65-66, No. 13). April 1966.

Five primary and three reserve equations were developed to span the quadrant containing all possible equations for three predictors, with the restriction that no weighcs be negative. Any empirical equation ever likely to be encountered operationally should be represented adequately by at least one of the primary theoretical composite equations with no more than .01 loss in predictive validity. The theoretical equations were positioned to achieve this result. It is not likely that additional equations will be needed; if so, the equations should be selected from the set of reserve equations to provide the most efficient coverage.

The theoretical system was applied to 47 prediction equations for different colleges using SAT-V, SAT-M, and high school rank in class as predictors and college performance as a criterion. The primary composite equations were selected as the appropriate composite equations for all of the original prediction equations except one. For that prediction equation, the appropriate composite equation is one of the reserve theoretical composite equations.

KASTRINOS, WILLIAM, AND GERTRUDE SANDERS. First-Year College Biology Courses: A Survey (RDR 74-75, No. 2). 1974.

At the request of the College Entrance Examination Board's Committee of Examiners in Biology, a questionnaire was developed to ascertain the kinds of first-year college biology courses that are being offered at colleges and universities in the United States. The committee sought to incorporate this information in the construction of a valid, representative course description upon which to base the Advanced Placement Examination in Biology.

The questionnaire was developed in the spring of 1972 and was sent to the biology department chairmen in the 100 colleges and universities receiving the highest number of requests for placement in biology.

Of the 100 colleges contacted, 68 completed the questionnaire and their responses comprise the data in this report. These data were made available to the Committee of Examiners to aid in the development of an appropriate course description and subsequently in the assembly of future forms of the Advanced Placement Examination in Biology.

KATZ, MARTIN R., LILA NORRIS, AND GERALD HALPERN. The Measurement of Academic Interests, Part I. Characteristics of the Academic Interest Measures (RDR 70-71, No. 4). October 1970.

Data presented in this study of the Academic Interest Measures (AIM) include score distributions for a national sample of high school juniors (N = 15,500), grouped by sex, ability, and curriculum (effects of socioeconomic status were found to be negligible); relationships between AIM and many other variables, such as ability test scores, students' ratings of their own interests, definiteness of occupational plans, years of postsecondary education planned, scores on information scales, and participation in nonclassroom activities; scale intercorrelations and factor analyses; and internal consistency and stability of scores.

The psychometric properties of AIM and evidence for its construct validity suggest that it may be useful as a criterion measure.

Data on predictive validities will appear in Part II.

Appendix D, consisting of the norms tables, is separately bound. Appendix E is a self-scoring form of AIM, devised by Paul B. Diederich.

LEVIN, MIRIAM K. The CEEB English Language-Study Test: A Survey of Linguistics Units in the High Schools (RDR 66-67, No. 7). December 1966.

Abstract not available.

LEVINE, MICHAEL V., AND DONALD B. RUBIN. Measuring the Appropriateness of Multiple-Choice Test Scores (RDR 76-77, No. 3). December 1976.

A student may be so atypical and unlike other students that his aptitude test score fails to be a completely appropriate measure of his relative ability. We consider the problem of using the student's pattern of multiple-choice aptitude test answers to decide whether his score is an appropriate ability measure. Several indicators of appropriateness are formulated and evaluated with a simulation of the Scholastic Aptitude Test.

LINN, ROBERT L. Grade Adjustments for Prediction of Academic Performance: A Review (RDR 64-65, No. 18). June 1965.

Several techniques of making grade adjustments are reviewed. Evaluations and comparisons of the various techniques are made in terms of the properties of the various models and on the basis of the empirical results based on the application of these models.

Although there are several relatively elegant models available, the empirical results have shown negligible improvement in predictive accuracy due to grade adjustments when (a) the results have been cross-validated and (b) one or more standardized test variables have been included as predictors.

LINN, ROBERT L., AND JUNIUS A. DAVIS. Correlates of Academic Performance of Community College Students in Career or Transfer Programs: A Pilot Study (RDR 66-67, No. 2). July 1966.

This study was undertaken (1) to investigate the utility of a number of promising experimental measures for predicting the performance and persistence of community college students and (2) to explore the goals and criteria for success within the various college programs. The major emphasis was on the career-oriented two-year programs.

Within the career programs, the Scholastic Aptitude Test and high school records had validities almost as good as they are at many four-year colleges. Several of the experimental measures showed promising validities within the career programs and/or distinctive patterns of means among the several program groups.

LINN, ROBERT L., DONALD A ROCK, AND T. ANNE CLEARY. Sequential Testing for Dichotomous Decisions (RDR 69-70, No. 3). May 1970.

Two sequential testing procedures for dichotomous decisions were investigated using existing item response data of 4,840 college students on three Achievement Tests. The assignment rules for the sequential test procedures and for short conventional tests were developed using half the sample of students and cross-validated using the other half. In general, the first type of sequential test required an average of approximately half as many items as the short conventional test for a given degree of accuracy. This is in close agreement with previous theoretical results. The second type of sequential test used knowledge of assignment on one dimension in making the assignment on the second method. The efficacy of this second method depends heavily on the relationship between the two dimensions.

MAIER, MILTON H., AND SCARVIA B. ANDERSON. Growth Study I. Adolescent Behavior and Interests (RDR 64-65, No. 6). October 1964.

The Background and Experience Questionnaire (BEQ) was developed to measure orientation of high school students toward either the adolescent or the adult culture. Both the type of behavior (TV viewing, reading, and activities) and the level of probable cognitive demands (high, medium, and low) were included in deriving the scores to measure orientations. Scores were also obtained to measure school interests, plans for the future, amount of study, part-time work, home background, and certain topics of thoughts and conversations. The BEQ was administered to a nationwide sample of high school seniors. Subsamples of 260 boys and 210 girls were used for a preliminary analysis, and the results in this report are based on these subsamples.

The high-level scores were positively correlated with measures of academic aptitude, socioeconomic status, interest in

academic school courses, and level of plans for the future. The low-level scores were either inversely related to or uncorrelated with these measures. The values of the correlations of the medium-level scores were in between those of the high- and low-level scores. The scores from the same level of probable cognitive demands had positive intercorrelations, whereas the high and low scores had either negative or near-zero correlations.

Predictions about the relationships between experiences outside of the classroom and measures of intellectual development are derived from an interactionist point of view of cognitive growth. The potential of the BEQ as an educational counseling tool is discussed.

MAJETIC, RICHARD M. The Characteristics of Preliminary Scholastic Aptitude Test User Schools and Non-User Schools (RDR 67-68, No. 2). July 1967.

Each year since 1959 the College Entrance Examination Board has offered the Preliminary Scholastic Aptitude Test for administration by schools. This testing program originated with three major purposes: (1) to offer an instrument that would give students information about themselves that would be helpful in planning their futures; (2) to assist schools in identifying potential college students who, not fully realizing their potential abilities, might not be planning to continue their education; and (3) to offer a less expensive and more timely test for students who would otherwise take the Scholastic Aptitude Test unnecessarily early and often in order to have their scores for making their college plans.

The number of schools that administer the PSAT and the number of students who take it attest superficially to its popularity, but these figures alone do not tell the extent to which the PSAT is now achieving its original purposes or any other useful purpose. Neither do they indicate how the program might be changed to serve such purposes more effectively.

A preliminary investigation has been undertaken to describe school characteristics of both "user" and "non-user" schools. These descriptions are based on demographic information readily available to the principals in a large number of schools. The use and availability of the PSAT has been investigated in the "user" schools as have the reasons for accepting the test in these schools and lack of use of the test by the "non-user" schools.

MARCO, GARY L. Item Characteristic Curve Solutions to Three Intractable Testing Problems (RDR 76-77, No. 5). June 1977.

This paper summarizes three studies that illustrate how application of the three-parameter logistic test model helped solve three relatively intractable testing problems. The three problems are designing a multi-purpose test, evaluating a multi-level test, and equating a test on the basis of pretest statistics. Examples use information about tests in various College Board testing programs. The first two studies demonstrate the value of the information function and relative efficiency. The last study is an application of basic item characteristic curve theory and shows the potential usefulness of pre-equating.

MILLS, DONALD. The Study of Change in CEEB Achievement Tests Over a Ten-Year Period (RDR 63-64, No. 3). November 1963.

Abstract not available.

MODU, CHRISTOPHER C. A Description of the Satisfaction Questionnaire for Junior Colleges in Terms of Rotated Factors (RDR 69-70, No. 1). March 1970.

This report describes the dimensions of a Satisfaction Questionnaire administered to junior college students during the experimental phases of the Comparative Guidance and Placement Program of the College Entrance Examination Board. Apart from investigating the stability of their factor structure, the scales were correlated with college grades to determine whether they measure something other than grades. The results produced relatively homogeneous scales indicative of the extent of a student's satisfaction or dissatisfaction with his English courses, mathematics courses, and major field and whether he had switched to a lower or more advanced section in English or mathematics. The scales were

also found to bear no relationship with college grades.

MODU, CHRISTOPHER C. The Effectiveness of an Essay Section in the American History and Social Studies Test (RDR 71-72, No. 6). February 1972.

The contribution of a 20-minute essay question, given as part of the one-hour Achievement Test in American History and Social Studies, to the pool of information available on a candidate from an all-objective examination of the College Board Admissions Testing Program is presented in this report. The study limits itself to a consideration of the psychometric issues. It does not deal with other important issues related to the educational impact of the essay or to the implications of the cost of introducing and maintaining a reliable essay section in the American History and Social Studies test. The results of the study led to the conclusion that the additional information gained by the introduction of an essay section is minimal. Moreover, the correlation of the essay with various subtests of the American History and Social Studies objective test suggests that the obtained essay scores may have reflected more of the factual content of the answers than the other skills.

MODU, CHRISTOPHER C., AND JUNE STERN. The Stability of the SAT Score Scale (RDR 74-75, No. 3). April 1975.

The purpose of this study is to determine the extent of drift that may have occurred in the SAT score scale between 1963 and 1973. In compliance with the recommendation in a 1961 report by Wilks to the College Board that a solid research effort be undertaken to assess a possible drift in the SAT verbal scale since 1941, a similar study, covering the period 1944 to 1963, was conducted by Stewart in 1966.

In the present study, raw scores on 1963 and 1966 forms of the SAT were placed on the reporting scale for a 1963 form. The raw-to-scaled score conversions derived in 1963 and 1966 for equivalent raw scores representing the same ability levels on these forms were compared with their corresponding scores on the 1973 reporting scale.

The equating procedures used in the experimental equating are identical with those used in the operational equating of SAT scores during the period of the study. However, in the operational equating, several equatings intervene between the tests which are equated directly in this study.

The results indicate that for both 1963 and 1966 equatings, the 1973 candidate group would have earned lower scores on both SAT verbal and SAT mathematical tests had the experimental equating results rather than the operational equating results been used in reporting the December, 1973, scores. The average difference is 14 points for SAT verbal and 17 points for SAT mathematical scores. The data of this study do not provide an explanation for the discrepancy between the results of the experimental and operational equating over the period of the study.

MOORE, MAXINE RUTH. A Survey of College Biology Departments Regarding Introductory Course Curricula and Advanced Placement Practices (RDR 63-64, No. 11). April 1964.

This study concerns an analysis of questionnaires completed by representatives of the College Entrance Examination Board Committee of Examiners for the Advanced Placement Examination in Biology on the basis of interviews with instructors of introductory biology courses in 32 colleges and universities throughout the nation. The questionnaires were concerned with the content of the introductory course and the experiences of the departments with entering students who took Advanced Placement Biology in high school. Opinions were also solicited regarding the correspondence of the Advanced Placement course description to the contents of beginning biology courses.

MYERS, ALBERT E., CAROLYN B. MCCONVILLE, AND WILLIAM E. COFFMAN. Reliability of Reading of the College Board English Composition Test, December 1963 (RDR 64-65, No. 3). August 1964.

In December, 1963, the College Board English Composition Test included a 20-minute essay. Approximately 80,000 candidates wrote the

examination, and the essays were read by 145 readers during the week following the administration. The purpose of this study was to assess the reliability of the reading.

From the total group of papers, 25 were randomly selected, reproduced by hand, and given to 25 different readers on each of the five days of the reading period. The experimental papers were mixed in with the mass of other essays so that the readers did not know there was a study in process or which papers were the experimental ones.

An analysis of variance procedure was used to assess the reliability of the judgments made by the readers. The reader reliability of the sum of two readings (the score used in reporting results) was .577, a figure that was consistent with those obtained in previous experimental investigations and indicated that there was no loss in reliability in moving from experimental reading of fewer than 1,000 papers to operational reading of more than 80,000 papers.

Single judgment reliabilities were .41 on the average, but this average was reduced by the very low .26 that was found for readings made on the last day. It was concluded that this drop probably represented an "end effect" in which the readers relaxed their vigilance as they approached the end of the demanding reading period. Other evidence suggested that the readers were making global judgments, as directed.

MYERS, CHARLES T. A Comparison of Three Methods for Scoring Multiple-Choice Tests (RDR 65-66, No. 6). March 1966.

This study compares the reliability and validity of scores obtained by using three different methods of scoring multiple-choice tests of English composition. The three methods were: the number of right answers, the number of right answers minus one-quarter the number of wrong answers, and the number of right answers minus the number of "worst" answers. The worst-answer key for each item was the one distracter that had been identified by expert judges as distinctively worse than the other distracters. Reliability was determined on the basis of split halves. The criterion for the validity for each of the three types of scores was based on a series of essay tests. For these tests the number-right score appeared to be slightly more reliable and slightly more valid than either of the other two scores.

NORRIS, LILA, AND MARTIN R. KATZ. The Measurement of Academic Interests, Part II. The Predictive Validities of Academic Interest Measures (RDR 70-71, No. 5). December 1970.

This study of predictive validities of academic interest measures is based on a follow-up through grade 12 and one year after high school graduation of students who had been tested in grade 11. The report focuses four major topics: (1) the prediction of marks in grades 12 and 13, with special attention to differential prediction; (2) the prediction of interests in grades 12 and 13, again emphasizing the contribution of academic interest measures to differential prediction; (3) the structure of abilities, interests, and values; and (4) the relationships between independent variables and occupational criteria for the subsample that did not continue in an educational institution during the year after graduation from high school.

Some incidental attention is also given to interest score profiles for various major-field and occupational groups. Another incidental analysis shows the proportions (classified by sex, socioeconomic status, and ability) of the students followed up after high school graduation who attended four-year or two-year colleges or did not go to college.

To assess the distinctive contribution to prediction made by the Academic Interest Measures (AIM), a consistent statistical procedure has been followed. Multiple correlations are given in a progression that always starts with Preliminary Scholastic Aptitude Test scores, verbal and mathematical (PSAT-V and PSAT-M), as predictors; then variables from AIM and other predictors are added successively. A major obligation is undertaken to determine the extent to which AIM increases predictive validities when ability scores, and sometimes previous marks, are already available. Finally, the contribution of AIM is compared with that of a simpler rating by students of their own interests in the fields represented by AIM scales.

PIKE, LEWIS W. Short-Term Instruction, Testwiseness, and the Scholastic Aptitude Test: A Literature Review with Research Recommendations (RDR 77-78, No. 2). January 1978.

The research literature on short-term instruction (STI) and intermediate-term instruction (ITI) for the SAT-M and SAT-V was reviewed. Selected studies of STI and ITI for tests other than the SAT-M and SAT-V, and of testwiseness (TW) per se, were included in the survey if they were judged relevant to the question of special instruction for the SAT.

The research studies were reviewed and interpreted within the framework of a score components model which posited four content-related and two TW score components, as well as two secondary ones (test-taking confidence and efficiency), that are theoretically subject to STI and ITI effects. In addition, examinee, item, and instructional characteristics were considered, as these relate to the score components model.

Basic discrepancies between negative and positive findings were noted for both the SAT-M and the SAT-V. These were generally resolved in favor of recognizing meaningful STI effects for the SAT-M, but remain unresolved for the SAT-V. Recommendations were made for SAT-M and SAT-V research allowing STI effects to be partitioned according to examinee, item, and instructional characteristics, as these apply to selected test score components.

PIKE, LEWIS W., AND FRANKLIN R. EVANS. The Effects of Special Instruction for Three Kinds of Mathematics Aptitude Items (RDR 71-72, No. 7). May 1972.

A different instructional program was developed for each of three mathematics aptitude item formats to determine the relative susceptibility of each to special instruction. Male and female high school junior volunteers in each of 12 schools were given a pretest composed of items of each format and a parallel form as a posttest several weeks later. In the intervening time, experimental subjects received seven weeks (21 hours) of instruction directed at one of the three formats, while control subjects received no special instruction.

Pretest to posttest gains were analyzed in a two-way (sex by instructional group) multivariate analysis of covariance. The six dependent variables were the geometry and nongeometry posttest scores for each of the item formats. The seven covariates included the pretest scores corresponding to the six dependent variables and the SAT verbal (SAT-V) pretest score.

Results of the statistical analysis showed that each of the three item formats was susceptible to the special instruction specifically directed toward it. The complex or novel item formats appeared to be more susceptible than the relatively straightforward item format. Female volunteers were found to be slightly less able mathematically at the outset and to benefit somewhat less from the instruction than male volunteers; mean gains of nearly a full standard deviation obtained by the groups instructed for the complex or novel formats were considered to be of practical consequence and likely to influence admissions decisions.

The results of the study were consistent for all 12 schools. Although no group received instruction for the SAT per se, substantial pre-to-posttest gains on that measure were also observed. Further analysis revealed that instruction, particularly for the complex or novel formats, was effective in at least two ways. Subjects appeared to have learned a systematic approach to the item format as well as some very basic mathematical concepts.

PITCHER, BARBARA. Follow-up Study of a National Sample of High School Seniors: Ability Level of Students Who Entered College Directly from High School and Were Graduated from College Within Four Years (RDR 68-69, No. 1). December 1968.

This study sought to answer the question: How do high school PSAT scores relate to college performance?

One of the findings was that the percentage of college entrants who were graduated within four years varied according to ability level (2V+M). Of the college entrants who scored in the top 5 percent of all high school seniors, 62 percent were graduated from college within four years. The corresponding percentage for the top 30 percent is 43; for those scoring below the 70th percentile, the percent graduated is 25.

This report should be of interest to administrators and planners in secondary and

higher education, high school counselors, college admissions and guidance personnel, researchers in secondary and higher education, and test development specialists.

The study follows an earlier study relating PSAT scores of graduating high school seniors to their records as college freshmen and hence strengthens the body of work dealing with the predictive validity of the PSAT. It should enable high school counselors and college admissions personnel to use PSAT scores to estimate with somewhat greater accuracy a student's probable performance in college. The broader social purpose is to help students find the college or vocation best suited to their abilities and interest.

PITCHER, BARBARA, AND JOHN SELSAM. A Preliminary Study of the Prediction of College Freshman English Grades from College Board Achievement Tests in the Foreign Languages (RDR 64-65, No. 4). September 1964.

Typically, in developing equations for predicting grades in college English, the researcher has used secondary school grades (or rank in class), Scholastic Aptitude Test verbal scores, and scores on the College Board English Composition Test. The question has been raised whether or not the accuracy of prediction would be improved if scores on foreign language tests were also used. This study, based on data available at Educational Testing Service, was undertaken to provide a tentative answer to the question.

The data come from four groups of college freshmen--two groups at men's liberal arts colleges and two groups at women's liberal arts colleges--who entered college in the fall of 1956 and for whom first-semester grades in English were available. All had taken the Scholastic Aptitude Test and College Board Achievement Tests in English Composition and in a foreign language (French, German, Latin, or Spanish) as candidates for admission to college.

The prediction of English grades from high school record, SAT verbal scores, English Composition Test scores, and foreign language test scores separately and in combination was studied for each of the four basic groups and for subgroups composed of students who had taken the French test and for subgroups who had taken a specific English course.

In general, the results indicate that, for these groups, the Achievement Test scores add little or nothing to the prediction of English grades from high school record and SAT verbal scores at the two men's liberal arts colleges. At the two liberal arts colleges for women, however, the language test scores as well as the English Composition Test scores look quite promising as predictors of English grades. In considering these data, it is extremely important to exercise caution and to think of these results as only preliminary answers to the questions with which this study began, for several reasons. First, the groups analyzed in this study are, in some cases, rather small portions of the total class admitted in fall, 1956, and, since they were all taking English courses in their freshman year and had all chosen the option of offering a foreign language test score for admissions purposes, they are probably more homogeneous than a total freshman class would be. Thus, although validity coefficients based on students already admitted to college are usually considered to be lower than they would be for unselected groups of candidates prior to admission, these coefficients are probably subject to this effect even more than usual. Second, the results found for these four northeastern College Board members in 1956 may not be representative of newer College Board members in other parts of the country or even of these same four colleges in 1964. It might be of interest to repeat the study for additional groups.

POWERS, DONALD E. Effects of "Taking the SAT" on Rate of Work and Guessing (RDR 80-81, No. 4). January 1981.

The College Board recently published Taking the SAT, a booklet designed to improve candidates' familiarity with the Scholastic Aptitude Test (SAT) and the Test of Standard Written English (TSWE). This note examines the extent to which receiving the new booklet altered (a) candidates' strategies with regard to responding in the face of uncertainty and (b) the rate at which examinees proceeded through the test. It is suggested that specific instructions found in the booklet are related to changes in examinees' test-taking behaviors in both of these areas.

POWERS, DONALD E. Students' Use of and Reactions to Alternative Methods of Preparing for the SAT (RDR 80-81, No. 3). December 1980.

A variety of courses, procedures, and materials are available to Scholastic Aptitude Test (SAT) candidates to help them prepare for the test. As potential consumers of these preparation resources, prospective examinees must decide which methods are most consonant with their own preparation objectives and which are likely to be effective. The sometimes conflicting claims and opinions of test sponsors, parents, and teachers, and of developers of test preparation resources, may pose a dilemma for students and their counselors.

This paper describes the incidence and patterns of use of various preparation resources in a large national sample of SAT candidates and presents examinee opinions regarding the usefulness of these resources. Data suggest that, although students exhibit a variety of attitudes about preparing for the SAT, most believe in the efficacy of some kind of preparation. This belief results in practical action, with most students preparing in some manner and a majority using more than one method of preparation. Although a large number of different styles were detected, a small number of patterns of use account for a substantial proportion of students' styles of preparation.

POWERS, DONALD E., AND DONALD L. ALDERMAN. The Use, Acceptance, and Impact of "Taking the SAT"--A Test Familiarization Booklet (RDR 78-79, No. 6). February 1979.

The College Board recently published Taking the SAT, a booklet designed to improve candidates' familiarity with the Scholastic Aptitude Test (SAT) and the Test of Standard Written English (TSWE), which is administered with the SAT. Prepublication copies were mailed to a random sample of SAT candidates, and the recipients, after taking the tests, were surveyed for their perceptions of the booklet and its influence on their test-taking. An equivalent random sample received only About the SAT, a briefer Board publication that Taking the SAT is to replace, and was surveyed for their test-taking experience and their opinions of the older test familiarization booklet. Test scores of the students in both groups were subsequently obtained from SAT score files for comparison.

The survey showed that the first group had reacted positively to the new publication and used it extensively. These candidates perceived numerous benefits from its use, some of them greater than the benefits perceived by candidates who had used only the older, briefer publication. The analysis did not, however, uncover any effects of the new booklet on scores for any section of the SAT or TSWE. Nevertheless, many students believed that Taking the SAT had helped them achieve higher scores.

The report documents the strong points of the new booklet and makes recommendations for improving some sections in future editions.

PRICE, DOROTHY BIRD. A Group Approach to the Analysis of Individual Differences in the Randomness of Guessing Behavior on Multiple-Choice Tests and the Development of Scoring Methods to Take Such Differences into Account (RDR 64-65, No. 10). December 1964.

Traditionally, formulas to adjust multiple-choice test scores for success from guessing have assumed that all guessing is purely random, although it is recognized that examinees attempt to narrow options.

The major purpose of this investigation was to see if groups of examinees, with similar ratios of right to wrong answers, differ sufficiently with respect to an analysis of the mean number of effective distracters (E) (a measure of the degree to which the number of item options were narrowed before guessing) that the E results for each group may be inferred to the individuals within that group.

The second purpose was to develop and study characteristics of scoring methods that take into account individual differences in narrowing behavior.

The data were sampled from two recent administrations of the Scholastic Aptitude Test--two verbal and two mathematical sections totaling 12,140 examinees.

Statistically significant variance analyses for all sections supported the conclusion that groups differentiated on the ratio of right to wrong answers (Ratio Groups) came from populations with different mean E values and that narrowing behavior increases as the right to wrong ratio

increases. From Ratio Group variance homogeneity tests, it was concluded that as the ratio increases, item differences in susceptibility to narrowing increase.

Differences among curves drawn through Ratio Group means were very small for all sections of the SAT studied. However, the differences led to the conclusion that tests of different content may be susceptible to different patterns of narrowing behavior. It was also suggested that a smoothed curve through mean E values for various Ratio Groups will intersect a similarly generated curve for various combined results of all the residual Ratio Groups, at a point at which examinees receive maximum benefit from narrowing, in relation to their knowledge.

The new scoring techniques used mean E values as best estimates for all test items. Curvilinear correlations between ratios and item E values indicated that the single estimate is a reasonably good predictor. However, an examination of the E curves for clusters of similar items showed that prediction curves for clusters of similar items could increase the importance of the ratio as a predictor of distracter effectiveness.

Four smaller samples from each SAT section were each scored in four ways. Method A corrected in direct relation to the amount of narrowing before guessing. Method B attempted to correct scores in accordance with the theoretically based amount that examinees benefit from narrowing in relation to their knowledge. The other methods were number of Right answers and Conventional Formula scoring.

The consistency of differences among intercorrelations of the scoring methods supported the conclusions that different tests of the same content and tests of different content display the same relationships among the four methods, and that in terms of ranking examinees differently from number of Right answers scoring, the order of increasing difference is from Conventional Formula to Method A to Method B.

The sixteen smaller samples were examined from the viewpoint of item analysis. Relationships between ratios and the two standard scoring methods left enough variance unexplained that it was concluded that grouping examinees by ratio rather than test score might yield a less-contaminated picture of option attractiveness. It was also concluded that ratio means and standard deviations could be valuable as additional criteria for parallel test forms.

Correlations between item E values and difficulty estimates supported the conclusion that in a well-written test, the relationship between item difficulty and number of effective distracters is very low.

For the samples studied, there were small but consistent reductions in reliability between number of Right answers and Conventional Formula and either Method A or Method B. It was concluded that correction formulas taking the ratio into account are only slightly less reliable than the Conventional Formula.

PRUZEK, ROBERT M., AND WILLIAM E. COFFMAN. A Factor Analysis of the Mathematical Sections of the Scholastic Aptitude Test (RDR 65-66, No. 10). April 1966.

The purpose of the research reported here was to explore ways of generating a factor pattern from interrelationships among items. The basic data consisted of tetrachoric correlations among the 60 items in the mathematical sections of one form of the Scholastic Aptitude Test. Two methods of factoring and four methods of analytical rotation were compared.

The factoring methods were: (1) a factoring of the tetrachoric correlation matrix with squared multiple correlations as communality estimates, the $R - S^2$ matrix, and (2) a principal axis factoring of the tetrachoric covariance matrix using communality estimates obtained by an iterative procedure. The rotations were the verimax, the equamax, the promax, and an oblique rotation recently described by Harris and Kaiser.

All methods produced compatible results; however, the clearest simple structure resulted when the first 11 factors extracted from the $R - S^2$ matrix were rotated to oblique simple structure by the oblique rotation described by Harris and Kaiser. Six factors were interpreted: geometric interpretation, computation, speed, arithmetic reasoning, data sufficiency, and rules.

RAGOSTA, MARJORIE. Handicapped Students and the SAT (RDR 80-81, No. 1). July 1980.

Abstract not available.

ROBERTS, S. O., AND DON B. OPPENHEIM. The Effect of Special Instruction Upon Test Performance of High School Students in Tennessee (RDR 66-67, No. 1). July 1966.

Previous studies of the effect of coaching on test scores have been essentially negative. The present study was undertaken to determine whether students who had received less adequate instruction and who were less able in scholastic aptitude test performance than the subjects in previous studies might benefit more from special instruction.

Specially prepared linear programmed materials, designed to foster an analytic approach to the tasks required on college admission tests and to be used in conjunction with classroom instruction, were developed in the verbal and the mathematical areas. Instruction in either of the two areas was given to eleventh grade students in 18 predominantly Negro high schools. Alternate forms of the PSAT, administered as pre- and posttests, were used to evaluate the effectiveness of the special instruction.

While there were statistically significant differences in gain scores between some of the experimental groups and their controls, the outcome of this study, like the outcome of earlier studies, was generally negative. In fact, the gains in the experimental groups were more than offset by reductions in posttest scores in the control groups. The magnitude of the gains observed in the experimental groups was so small that it does not seem reasonable to expect that similar short-term instruction given on a wide scale would be of significant benefit to disadvantaged students.

ROCK, DONALD A., AND CHARLES E. WERTS. Construct Validity of the SAT Across Populations--An Empirical Confirmatory Study (RDR 78-79, No. 5). April 1979.

An analysis of covariance structural factor model was posed that allowed for tests of invariance across populations of (1) similarity of the factor pattern, (2) equivalence of units of measurement, (3) standard error of measurement of each factor, (4) reliability of each factor, and (5) a factor mean structure. The analysis of the SAT-V and the SAT-M indicated support for the hypothesis that the respective factors for SAT-V and SAT-M were measuring the same things across populations in the same units with equivalent accuracy as measured by the standard error of measurement.

The factor analysis of item type classifications indicated that little useful information would be obtained if the scores were reported by item type. This is particularly true for the SAT-M. The verbal item type analysis did suggest the possibility that the analogy scores be reported separately and the remaining item types be combined in a single reading-vocabulary score.

ROCK, DONALD A., LEONARD L. BAIRD, AND ROBERT L. LINN. Interaction Between College Effects and Students' Aptitudes (RDR 70-71, No. 7). April 1971.

The effects of college environments on student achievement on three GRE area achievement tests, when the initial ability of students was controlled, were studied. A wide variety of environmental measures for 95 colleges were included. The method computed within-school regression lines, then colleges were clustered on regression lines, using a taxonomic technique, and multiple group discriminant functions using college descriptive variables were computed.

It was found that in all three achievement areas, one group or cluster of colleges could be identified as being more effective than the remaining groups of colleges. In the humanities areas the most effective group was characterized by a higher proportion of faculty with doctorates, greater selectivity, and a larger budget. In social sciences and the natural sciences, emphasis on subject-matter area as measured by the proportion of majors in the appropriate fields of study was the only variable that demonstrated some discrimination between the most and least effective groups.

ROSS, DONALD C. High School Curriculum and Performance on Mathematics Achievement Tests (RDR 67-68, No. 3). January 1968.

Two new Mathematics Achievement Tests were devised. The Level II test contains items that are more difficult than those that are found on the Level I test. It also contains items that measure areas that are more likely to be covered in courses that have a contemporary emphasis as opposed to courses

that have a more traditional emphasis. It was found that among a group composed largely of engineering students, those students who elected to take the Level II test showed a higher mean level of performance on the SAT-M and the Advanced Mathematics Achievement Test than those who elected to take the Level I test. They also reported different emphases in the secondary school mathematics courses they took. It was found that although there were correlations between test performance and measures of the type of secondary school courses, with respect both to content covered and to level of preparation, these correlations were no greater for either of the new tests than for the earlier Advanced Mathematics Achievement Test.

SCHLEKAT, GEORGE A. Financial Aid Awards of High School Seniors in the Northeastern States as a Function of Socioeconomic Class (RDR 67-68, No. 4). March 1968.

The study analyzes the relationship between socioeconomic status of financial aid applicants and college practices in distributing aid, taking account of measured academic ability and financial need.

Five socioeconomic classes are distinguished, ranging from upper to lower; colleges are grouped according to source of control, type of student body, and type of institution; and awards are classified in seven combinations of grant, loan, and job. The study is limited to the northeastern area of the United States, where the Scholastic Aptitude Test is predominantly used as a basis for college admission and financial aid decisions.

Socioeconomic status was determined from variables extracted from approximately 16,000 Parents' Confidential Statements, randomly selected from files accumulated during the 1964-65 academic year. One thousand colleges returned 19,131 usable questionnaires. Award offers were made to 6,579 of the applications represented by the questionnaires and were included in the analysis.

Results of analyses of variance and covariance procedures indicate that there is no straightforward answer to the question of whether socioeconomic classes are treated differently. In some ways upper classes receive preferential treatment, apparently because of superior test scores; lower classes are treated better in other ways, partly because of their poverty.

It may be said that lower-class applicants have a better chance of receiving aid, and in larger amounts, than their upper-class peers. But their awards more often consist of loans and jobs rather than outright grants which are more frequently awarded to the upper-class applicant. Thus, the upper socioeconomic classes are more likely to graduate from college with less indebtedness than lower classes, who not only mortgage their future more heavily but are required to devote more of their free time to gainful employment.

SEIBEL, DEAN W. Follow-up Study of a National Sample of High School Seniors: Phase 2--One Year After Graduation (RDR 65-66, No. 1). October 1965.

A follow-up study of a national sample of high school seniors was conducted to obtain information regarding the students' activities during the first year after graduation from high school. Information was obtained on (1) whether the student enrolled in a college and (2) how he performed in the first year of college if he enrolled. This information is related to measures of scholastic aptitude and performance obtained while the student was a senior in high school.

The report presents data for the Preliminary Scholastic Aptitude Test, the Scholastic Aptitude Test, and high school rank in class for a number of criterion groups based on college attendance and performance, and examines the relationships among PSAT, high school performance, and college performance.

SEQUEIRA, LEFORNE. A Study of the CGP Financial Need Indicator (RDR 70-71, No. 3). October 1970.

The information obtained from students on the CGP/BIB is adequate within the program's definitions to identify students who may need special counseling regarding financial assistance. The FNI, designed to initiate counseling activity by describing relative need in gross terms within broad categories of college costs, is fulfilling its goal, even when the information supplied by

students may not be highly accurate in terms of specific amounts the parents may report.

Student-supplied information reported more accurately would, of course, add to the value of the instrument. Revisions to the 1970-71 CGP/BIB include more income levels that a student may identify, narrower ranges of income within item responses, and an additional college cost category. The revisions should result in more refined identification within the FNI.

The evidence also suggests that the technique used in the FNI be investigated for use in programs other than CGP and also that further investigation be carried out to ascertain its value in more sophisticated need analysis instruments.

SERLING, ALBERT N. Three Methods of Testing Literary Comprehension at the Advanced Placement Level (RDR 66-67, No. 4). March 1967.

This study was intended to assess the reliability and validity of essay, short-answer, and multiple-choice questions as measures of the ability of Advanced Placement (AP) candidates to understand, analyze, and interpret literature. Before 1966, AP Examinations in English had consisted of essay and short-answer questions only: the reliability of these, particularly the essay questions, had been lower than is desirable. However, the Committee of Examiners had been reluctant to alter the composition of the test in the absence of evidence that change would be beneficial and that a test composed in part of objective questions would provide a more reliable measure of the ability to understand and interpret literature at no real sacrifice of validity.

SHARON, AMIEL T. Effectiveness of Remediation in Junior College (RDR 70-71, No. 2). September 1970.

Randomly selected groups of students required to take remedial courses at two community colleges were placed in regular courses on an experimental basis. The experimentally placed students were compared with those who enrolled in a regular course after passing a remedial course and with those who did not require remediation. Comparisons were made on ability, interest, motivation, persistence, and performance. The English remedial course had no effect on student satisfaction but produced a small improvement in subsequent performance. The mathematics remedial course eliminated some of the dissatisfaction with the regular course and had a significant effect on subsequent course work. The placement procedures appeared to be more effective in assigning students to appropriate mathematics than to English courses.

SHARON, AMIEL T. Measurement of College Achievement by the College Level Examination Program (RDR 69-70, No. 5). June 1970.

The criterion-related and construct validity of the College Level Examination Program's General Examinations are discussed primarily in terms of research studies conducted at institutions of higher education. While most of the research provides support for the validity of the examinations as measures of academic achievement in college, the results of many of the studies have alternative explanations. The examinations correlate positively with course grades and amount of previous college instruction. Significant gains on the tests are generally made by students over the first two years of college and the highest scores on each test are obtained by students intending to major in the subject covered by the test. There is some question as to whether each of the five examinations is measuring a unique factor.

STATISTICAL ANALYSIS DIVISION. Using Scholastic Aptitude Test Scores in Counseling Prospective Graduate and Law Students (RDR 63-64, No. 6). October 1963.

During his years of secondary school and college, a student accumulates information which is highly relevant to his future educational plans. By the end of his freshman year of college, for example, he should know a great deal about his ability to do college work and about how well he can discipline himself to meet complex and exacting tasks. He should have at least a preliminary idea of the direction of his interests, and he may have clearly defined goals. His knowledge of the demands and opportunities offered by various fields of st dy and work

should have progressed beyond a superficial familiarity. The student who thinks seriously and uses available information well will enhance his chances of making wise decisions as his career takes shape.

Clearly, test scores should contribute to the student's self-understanding. Their usefulness for planning by undergraduates is limited, however, by the fact that different tests are used for graduate and professional school admissions from those used by colleges. Thus, these tests have tended to be an unknown quantity in educational planning until they are taken, late in the student's college career. Fortunately, scores on some of these tests can be predicted from College Board Scholastic Aptitude Test (SAT) scores almost as well as four-year college average grades can be predicted from freshman averages. Data on these relationships are presented in the belief that they will help the undergraduate in his educational planning and will reduce somewhat the air of mystery which sometimes surrounds the tests.

STEWART, E. ELIZABETH. The Stability of the SAT Verbal Score Scale (RDR 66-67, No. 3). July 1966.

In his 1961 report to the College Board, Wilks recommended that "The College Board should conduct a solid research effort to make an assessment of the extent of drift which may have occurred in the SAT verbal scale since 1941." In the present study, which is based on a design suggested by Angoff, raw scores on each of four old forms of SAT-V were placed on the reporting scale for a recent form so that the resulting scaled scores could be compared with the scaled scores derived when the old form was first administered in the College Board Admissions Testing Program. The data on the recent form were collected as part of the December, 1963, administration of the Scholastic Aptitude Test; data on the old forms were retrieved from files for the April, 1944, January, 1948, March, 1953, and February, 1957, SAT administrations. For each of these forms, the scale derived by equating scores on the old form to scores on the December, 1963, form was compared with the scale used for reporting scores on the old form.

The results may be summarized as follows:

(1) For both the 1953 form and the 1957 form, the experimentally derived scales and the reporting scales differed by less than five points in the vicinity of the mean. The degree of agreement was good enough to warrant the conclusion that the SAT verbal scale remained highly stable during the period from 1953 to 1963.

(2) With the reservation that the standard errors of both the operational and the experimentally derived scales for the January, 1948, form probably are relatively large, it may be concluded that the SAT verbal score of a candidate tested in 1963 and scoring in the vicinity of 500 might be 20-35 points higher than the score of a candidate of equal ability who was tested in 1948. Since the scale as it existed in 1953 was maintained through 1963, it may be inferred that the shift in the meaning of the scale took place between 1948 and 1953.

(3) Scores on SAT-V reported in April, 1944, may not have had the same meaning as scores reported in December, 1963, but the methodological limitations inherent in the experimental equating of those two forms preclude any reliable quantitative assessment of whatever differences may exist between the scale of the early 1940s and the scale of the early 1960s.

STICKELL, DAVID W. A Comparison of the Performance of BSCS and Non-BSCS Candidates on the CEEB Biology Achievement Test May 1963 Administration (RDR 64-65, No. 12). March 1965.

From data collected on candidates who took the College Board Biology Achievement Test in May, 1963, analyses were carried out to test the significance of differences between the adjusted mean score on biology of non-BSCS, or "conventional," students and each of three other groups of students who had taken one of the courses prepared by the Biological Sciences Curriculum Study: (1) the Blue version, (2) the Green version, and (3) the Yellow version. Analyses were based on students who took the Biology Achievement Test in May, 1963; Biology scores were adjusted for differences on SAT-V, on SAT-M, and on a subset of biology questions that were judged to be appropriate for both the students who took a "conventional" biology course and those who took a particular BSCS course.

After adjustment, there was a difference of 18 points on the College Board scale in favor of the Yellow version students over "conventional" students. There were differences of 12 and 10 points in favor of the "conventional" students over the Blue and Green version students, respectively. Only the difference in favor of the Yellow version students was significant at the .05 level.

STICKELL, DAVID W. An Analysis of Culver Military Academy Data Relating to Performance on the College Board Chemistry Test as Affected by Curricular Change (RDR 64-65, No. 17). June 1965.

In the fall of 1961, Culver Military Academy began teaching the CHEM Study course to all students taking chemistry. At the previous May administration of the College Board Chemistry Test, Culver students earned a mean score of 553; in May, 1962, the mean Chemistry score of Culver students dropped to 481. The Academy was naturally concerned about this abrupt change in the scores of their chemistry students and corresponded with Educational Testing Service. Out of this correspondence came the decision to make a more intensive study.

With data pertaining to Culver chemistry students from 1960 through 1963, two analyses of covariance were carried out. One analysis tested the significance of differences among mean chemistry scores for the four years, adjusted for differences on verbal and mathematical Scholastic Aptitude Test scores. The differences were significant and t tests were calculated to test the significance of differences between 1960 and 1961 means, between 1962 and 1963 means, and between the combined 1960 and 1961 means and the combined 1962 and 1963 means. The first two differences were significant at the .05 level, and the third difference was significant at the .001 level.

The second analysis, for comparison purposes, tested the significance of differences among mean English Composition Test scores for the four years, adjusted in the same way as were the Chemistry means. The results were not significant at the .05 level.

The report ends with a discussion of some of the problems encountered in determining the relative appropriateness of a College Board Achievement Test for students of different courses.

TEMP, GEORGE. Test Bias: Validity of the SAT for Blacks and Whites in Thirteen Integrated Institutions (RDR 70-71, No. 6). January 1971.

Differential prediction for Black and white students was empirically investigated at 13 institutions by comparison of regression planes. Particular attention was given to the possibility that prediction procedures that are appropriate for white (majority) students would underpredict the performance of Black (minority) students. The data tend to support, among others, the following generalizations: (1) a single regression plane cannot be used to predict freshman GPA for both Blacks and whites in many of the institutions studied; (2) nevertheless, if prediction of GPA from SAT scores is based upon prediction equations suitable for majority students, then Black students, as a group, are predicted to do about as well as (or better than) they actually do.

THOMPSON, RAYMOND E. A Study of the Comparative Predictive Validities of the Essay and Objective Sections of the College Entrance Examination Board Advanced Placement Examination in Physics (RDR 65-66, No. 2). October 1965.

The comparative predictive validities of the one-hour objective section and the two-hour essay section of the College Entrance Examination Board Advanced Placement Test in Physics were investigated using grades in college physics courses as the criterion. The study included 226 students who took one form of the test in 1962 and then earned grades in one or more of 57 courses and 222 students who took another form of the test in 1963 and then earned grades in one or more of 33 courses. Between 5 and 48 examinees earned grades in each of the 90 courses considered.

Two methods of correlational analysis indicated that the scores on the objective and essay sections of the test were about equally effective for predicting grades in college physics courses. For the 57 courses in which students tested in 1962 earned grades, the mean correlation between

objective scores and grades was .34, between essay scores and grades .38, and between composite scores and grades .40. For the 33 courses in which students tested in 1963 earned grades, the mean correlation between objective scores and grades was .35, between essay scores and grades .26, and between composite scores and grades .33. For all students tested in 1962 combined in one group without regard to the colleges they attended or the courses they took, the correlation between objective scores and mean course grades was .27, between essay scores and mean course grades .26, and between actual composite scores and mean course grades .29. For all students tested in 1963 combined in one group, the correlation between objective scores and mean course grades was .30, between essay scores and mean course grades .23, and between actual composite scores and mean course grades .28. For these same two groups of students tested in 1962 and 1963, the correlations between composite scores derived from the most optimally weighted combination of objective and essay scores and mean course grades would have been .29 and .30, respectively. These figures represent only a trifling improvement over the correlations obtained between the actual composite scores and mean course grades despite the fact that the actual composite scores were derived using weights that differed widely from the optimal weights.

Correlations between test scores and course grades for examinees in different courses varied greatly, ranging from -.80 to .95. An analysis of mean correlations between test scores and course grades for groups of courses categorized by type (introductory, advanced laboratory) yielded no evidence that the test or one of its sections was a better predictor of grades in one type of course than in another.

THOMPSON, RAYMOND E. Developing a Collegiate Physical Science Test (RDR 63-64, No. 4). December 1963.

Abstract not available.

THOMPSON, RAYMOND E. Investigations of the Appropriateness of the College Board Science Achievement Tests for Students of Different High School Science Courses (RDR 71-72, No. 2). September 1971.

The appropriateness of the College Board Achievement Tests in Biology, Chemistry, and Physics for students of different high school science courses was investigated. These investigations involved six Biology, twelve Chemistry, and eight Physics Achievement Test forms introduced in the 1960s. The comparative fairness of the tests for various pairs of courses was analyzed. These pairs were: Biological Sciences Curriculum Study-Blue and Regular Biology, BSCS-Green and Regular Biology, and BSCS-Yellow and Regular Biology; Chemical Bond Approach Project and Regular Chemistry and Chemical Education Material Study and Regular Chemistry; and Physical Science Study Committee and Regular Physics.

For most test forms, three indications of their appropriateness for the courses in each relevant pair were secured. The first was ratings of test questions by teachers of the two courses. The second indication came from the actual performance of students of the two courses on the complete science Achievement Test, on questions in the tests rated appropriate for both courses by teachers, on the College Board Scholastic Aptitude Test verbal section, and on SAT mathematical section. The third indication came from an analysis of the Achievement Test scores of students of the two courses after adjusting for their differing performances on the concomitant measures of the appropriate-for-both questions, SAT-V, and SAT-M.

In interpreting the three indications of appropriateness, the following standards were taken to show bias: (1) mean teacher ratings differing significantly at the .05 level; (2) actual means on the Achievement Tests differing in one direction with actual means on the concomitant measures differing in the opposite direction; and (3) adjusted Achievement Test means differing by more than 15 scale score points after controlling for performance on the concomitant measures through the analysis of covariance.

WARREN, JONATHAN R. Prediction of College Achievement Among Mexican-American Students in California (RDR 76-77, No. 1). July 1976.

Equations for the prediction of college grades from test scores and high school grades were compared for Mexican-Americans and other students at three universities of the California State University and Colleges System. With rare exceptions, regression equations for the prediction of grades for Mexican-Americans did not differ from those for other students at the freshman, sophomore, or junior level. The prediction of senior-year grades did require different equations for Mexican-Americans, but the direction of the difference was reversed at the two universities involved. The need for separate prediction equations for Mexican-American applicants to college is not supported. Nevertheless, the modest size of the multiple correlation coefficients, usually in the range from .30 to .50, suggests that considerations other than test scores and prior grades should enter admissions decisions. The few comparisons across universities did not indicate a need for different equations for different institutions, but a difference in grade distributions at the two institutions compared suggests that the question be examined again with more substantial data. Men and women differed in their prediction equations, although neither sex was consistently over- or underpredicted. Men, however, almost invariably showed higher test scores and lower grades at all levels than did women.

WARREN, JONATHAN R., AND PAMELA J. ROELFS. Student Reactions to College: The Development of a Questionnaire Through Which Junior College Students Describe Their College Experiences (RDR 72-73, No. 3). October 1972.

A large number of interviews were conducted with students and staff of junior colleges across the country. Analysis and discussion of the interview results at two conferences of junior college experts provided guidelines for development of a questionnaire that would provide the desired vehicle for students to express their views of their junior college experiences. Items were written reflecting the content and emphasis brought out in the interviews. Several tentative forms for presenting and organizing the items were attempted and abandoned before the trial version was produced for a widespread administration.

A form consisting of 203 items plus provisions for write-in items and comments was completed in the late spring of 1971 by 6,495 students at 27 junior colleges. Examination of the response distributions, further review by an advisory group, and incorporation of some of the recurrent write-in comments led to item deletions, revisions, and additions. The resulting questionnaire consists of 157 items.

WATKINS, RICHARD W., AND RICHARD S. LEVINE. The Usefulness of Adjustments to Secondary School Grades in the Prediction of College Success (RDR 68-69, No. 2). February 1969.

This report presents the results of two studies conducted for the College Entrance Examination Board by Educational Testing Service. The studies were directed to the question: Can the accuracy of predicting college grades be improved by differential adjustment of measures of high school achievement which take into account differences in college achievement of graduates of different high schools? In each study, two methods of adjusting grades were applied.

The first study, a pilot study for the second and larger one, utilized data provided by some colleges and permitted the adjustment of class ranks from 45 secondary schools. Scores derived from the class ranks and size of the class were adjusted by two methods that differed in the amount and kinds of information used. The first method used only the SAT verbal scores obtained by graduates of the high schools for which the rank scores were to be scaled. The second method utilized SAT verbal scores to place the grade averages for the seven colleges on comparable scales, and then scaled the high school rank scores on the basis of college grades on the comparable scales obtained by the high school graduates.

The median increases in the correlation coefficients obtained for the two methods of adjustment in comparison with those obtained using the reported class ranks at the seven colleges were .04 and .06. When the reported or adjusted class ranks were combined with the SAT scores to obtain multiple correlation coefficients, the median differences in the correlation

coefficients as a result of the scaling were reduced to .02 and .03. Finally, the use of both SAT scores and the average score on three College Board Achievement Tests in conjunction with the reported rank scores yielded higher correlation coefficients than did either of the adjusted rank scores used in combination with the SAT scores. In other words, the addition of Board Achievement Tests to the SAT and reported rank scores is a more effective way of increasing the correlation of preadmission information with college grades than is the adjustment of the reported rank scores.

The second study was intended to be as much a study of the operational feasibility of a large-scale grade adjustment system as it was a study of the effectiveness of grade adjustment per se. In it, the relative usefulness of adjusting each one of a variety of high school grade averages and rank were compared; in addition, the effectiveness of any adjustment of high school achievement measures was investigated. The study used data provided by 14 colleges and universities and 186 high schools in the state of Indiana.

The first method of adjustment was carried out using appropriate SAT scores (verbal or mathematical or the sum of the two) to adjust eight different measures of high school achievement. The results for only four of these eight measures are discussed in this report since the results were essentially the same for all eight measures. The second method of adjustment was one designed to yield a fairly good estimate of the maximum increase in the correlation coefficients that could be obtained as a result of any adjustment process.

The results of the grade or rank adjustments were essentially similar to those obtained in the pilot study, although the medians of the increases in the zero order correlation coefficients (.02, .06) tended to be smaller. For the colleges, the multiple correlation coefficients of test scores and reported grades or ranks were almost as high as those obtained when adjusted grades were used. For specific colleges, the use of the reported grades or rank yielded more accurate predictions because the operationally feasible method of grade adjustment reduced the accuracy of prediction for students in some of the college groups.

These results, and those from a number of other studies, indicate that the introduction of a large-scale system for adjusting measures of high school achievement will not result in an improvement in the accuracy of prediction over what can be obtained with much simpler and less expensive prediction systems.

WILSON, KENNETH M. Multivariate Talent Flow Analysis: A Pilot Study (RDR 78-79, No. 3). October 1978.

This heuristically-oriented study employed a wide range of data from the College Board Admissions Testing Program file on college-bound students to explore a multivariate approach to talent-flow analysis, or analysis of differences between admissions-stage groups. Multivariate talent-flow analysis, a largely undeveloped area of admissions-related research, may be thought of as having two related general objectives:

(a) to analyze the correlates of dichotomous flow criteria--i.e., to determine which of several student variables, or talent descriptors, are associated with filing an application for admission (the flow criterion is application versus nonapplication), being offered admission (accept vs. reject), and enrolling (enrolling vs. not enrolling), respectively; and to assess the relative importance of the student variables as factors affecting, or involved directly or indirectly in, advancement/nonadvancement at each admissions stage; and

(b) to predict the respective flow criteria--i.e., to determine rates or probabilities of application, acceptance, and enrollment, respectively, according to score-level on composites of relevant student variables optimally weighted to discriminate between criterion groups at each admissions stage.

Results of the exploratory study, using data for college-bound students sending CB score reports to one state university, strongly suggest that the type of information yielded by using the multiple regression or discriminant model to analyze differences between admissions-stage groups with respect to a variety of academic, personal, and demographic data (talent descriptors) should be of value to colleges and universities interested in assessing and evaluating their admissions and recruitment policies and procedures. Moreover, the optimally weighted discriminant composite of variables found to differentiate admissions-

stage groups provides information that can be used to facilitate prediction of yield--e.g., to determine which candidates are most likely to complete an application or to enroll if offered admission.

From the point of view of accountability in admissions, systematic analysis of the correlates of the accept/reject decision would appear to be relevant. By regressing the accept/reject criterion on appropriate student variables, it would be possible to ascertain the relative role in admissions of, say, academic variables (such as tested ability and previous academic record) and nonacademic and/or socially relevant variables (such as, for example, socioeconomic status, sex, minority vs. nonminority status).

Knowledge of the correlates of individual decisions to apply or to enroll if accepted should be useful to institutions interested in identifying high-yield candidate categories in connection with their recruitment programs.

Analysis of student data from a variety of colleges and universities representing different admissions modes and circumstances is needed to establish the extent and nature of variation in the patterns and the correlates of application, admission, and enrollment according to institutional differences in admissions policies and emphases.

WILSON, KENNETH M. Predicting the Long-Term Performance in College of Minority and Nonminority Students: A Comparative Analysis in Two Collegiate Settings (RDR 77-78, No. 3). April 1978.

In two undergraduate settings, data reflecting the cumulative records of cohorts of minority and nonminority students over from four to seven years following matriculation were analyzed to determine the correlational validity of admissions tests and school rank for predicting long-term cumulative GPA and non-GPA criteria (e.g., degree attainment), as well as short-term, first-year GPA, the criterion most often used in comparative validity studies. An analysis of trends in mean GPA from the first through the eighth semester and cumulatively was undertaken to explore the possibility of "late blooming" tendencies among minority students--i.e., the possibility that minority students may show relatively greater improvement in academic performance following the first year than their nonminority classmates.

The late-bloomer hypothesis derives from the plausible argument that minority students face special problems of school-to-college transition that adversely affect their first-year performance. After a period of adjustment, the argument continues, tendencies to perform at a first-year level no higher and frequently lower than expected from test scores may be halted or even reversed.

Data employed in the study were from two quite different settings; namely, Main Campus (of a complex state university system) and College (a selective, liberal arts college for men). Results of correlational analyses in both settings indicated that the admissions variables used were at least as valid for predicting longer term criteria of success (e.g., overall or four-year cumulative GPA, highest educational level attained) as they were from predicting short-term first-year GPA, and that this finding was consistent for minority as well as for nonminority students.

With respect to the novel question of late blooming, GPA trends at Main Campus were not consistent with the late-bloomer hypothesis--i.e., minority and nonminority students were somewhat more sharply differentiated by later semester and overall cumulative GPA than they were by the first-year or first-semester GPA. At College, which provided data for four successive entering classes (cohorts), results of GPA analysis were ambiguous, suggesting the possibility of late-blooming tendencies in data for later but not for earlier cohorts.

Interpretation was complicated by substantial across-cohort (across-class) increases in average grades, greater for minority than for nonminority students, that could not be explained by increases in ability levels (test-score averages did not increase across classes). "Inflationary" and "emergent late blooming" rationales were offered for these findings, but they could not be rigorously evaluated due to lack of needed data.

According to the inflationary rationale, the extra across-cohort GPA increases for minority students could be accounted for by changes in the comparability of faculty grading standards and performance expectations for minority and nonminority students and shifts among minority students toward "less demanding" courses or curricula, etc. According to an emergent late-blooming

rationale, the extra minority GPA gains could be due to possible increases across successive entering cohorts, plausibly greater for minority students, in average levels of academic and social sophistication, general coping ability, self-confidence, etc. Such increases would not necessarily be reflected in test-score averages, but if present could be conducive to improved levels of academic productivity; improved institutional support systems for minority students could also be involved.

The concept of "emergent late blooming" focuses attention on the possibility of generational, developmental increases among minority students in average levels of academic intellectual and social sophistication, self-confidence, achievement motivation, and other noncognitive performance-related characteristics--increases that might reasonably be expected to occur over time as the more obvious barriers to full equality of opportunity for minorities are removed. The unexplained across-cohort increases in average grades and especially the interpretive problems introduced by differential rates of increase for minority and nonminority students point up limitations of the grade-point average as an index of academic performance in comparative validity studies generally that should be given special consideration in future research concerned with across-cohort developmental trends in the comparative performance of minority and nonminority students.

INDEX OF AUTHORS

A

Alderman, Donald L., 1, 29
Anderson, Scarvia B., 23
Angoff, William H., 1, 2

B

Baird, Leonard L., 31
Boldt, Robert F., 3
Breland, Hunter M., 3, 4
Bridgeman, Brent, 5

C

Campbell, Joel, 5
Cardall, Carolyn, 5
Casserly, Patricia L., 6, 7
Centra, John A., 7, 8
Cleary, T. Anne, 8, 9, 23
Cliff, Norman, 9
Cliff, Rosemary, 9
Coffman, William E., 5, 7, 9, 10, 11, 14, 15, 25, 30

D

Davis, Junius A., 11, 12, 23

E

Ekstrom, Ruth B., 12
Evans, Franklin R., 13, 27

F

Flaugher, Ronald L., 13, 14
Ford, Susan F., 1
Fremer, John, 14
French, John, 15

G

Gitlitz, Alfred H., 15
Godshalk, Fred I., 15
Grandy, Jerilee E., 16

H

Halpern, Gerald, 16, 17, 22
Harrison, Sylvia E., 17
Hartnett, Rodney T., 17
Haven, Elizabeth W., 18
Heidgerd, Lloyd H., 18
Hilton, Thomas L., 9, 18
Horch, Dwight H., 19
Howell, John J., 19, 20

J

Jackson, Rex, 20
Jacobs, Paul I., 21
Jones, Chancey O., 21

K

Karas, Shawky F., 21, 22
Kastrinos, William, 22
Katz, Martin R., 22, 26
Kaufman, Nadeen L., 15
Kendall, Lorne M., 22
Kline, Roberta G., 5
Kurfman, Dana G., 10

L

Levin, Miriam K., 22
Levine, Michael V., 23
Levine, Richard S., 37
Linn, Robert L., 23, 31

M

Mahoney, Margaret H., 10
Maier, Milton H., 23
Majetic, Richard M., 24
Marco, Gary L., 24
McConville, Carolyn B., 25
Mills, Donald, 24
Minsky, Shula, 4
Modu, Christopher C., 1, 24, 25
Moore, Maxine Ruth, 25
Myers, Albert E., 18, 25
Myers, Charles T., 26

N

Neun, Mary Ellen, 11
Norris, Lila, 22, 26

O

Oppenheim, Don B., 31

P

Peterson, Richard E., 7
Pike, Lewis W., 14, 27
Pitcher, Barbara, 27, 28
Powers, Donald E., 1, 28, 29
Price, Dorothy Bird, 29
Pruzek, Robert M., 30

R

Ragosta, Marjorie, 30
Roberts, S. O., 31
Rock, Donald A., 14, 23, 31
Roelfs, Pamela J., 37
Ross, Donald C., 31
Rubin, Donald B., 23

S

Sanders, Gertrude, 22
Sasajima, Masu, 16, 17
Schlekat, George A., 32
Schrader, William B., 20
Seibel, Dean W., 32
Selsam, John, 28
Sequeira, Leforne, 32
Serling, Albert N., 33
Sharon, Amiel T., 2, 19, 33
Smith, Robert E., 18
Statistical Analysis Division, 33
Stern, June, 2, 25
Stewart, E. Elizabeth, 34
Stickell, David W., 34, 35
Swineford, Frances, 15

T

Taylor, Phillip H., 14
Temp, George, 35
Thompson, Raymond E., 35, 36

W

Warren, Jonathan R., 37
Watkins, Richard W., 37
Werts, Charles E., 31
Williams, S. Irene, 21
Wilson, Kenneth M., 38, 39